Using the Landsat Catalogue to Assess Thirty Years of Change in Native Woody Vegetation in Three New Zealand Sheep and Beef Farming Regions

Joshua Foster

A thesis presented for partial fulfilment of the degree of

Master of Forestry Science

At the

School of Forestry

University of Canterbury

Christchurch New Zealand

31st of May 2020

SUPERVISORS:

Prof. David Norton
Dr. Bradley Case
Dr. Cong Xu

Project Summary

The aim of this research was to produce a time series analysis of change in native woody vegetation in selected sheep and beef farming regions using satellite imagery attained from the Landsat programme. This was proposed to be achieved by training a machine learning classifier to accurately estimate vegetation attributes in modern imagery and applying the produced models to historic imagery that predates the New Zealand Land Cover Database. Training machine learning classifiers to recognise the desired suite of land cover classes in modern satellite imagery proved to be a challenging task in itself, and due to image incomparability issues, applying these trained classifiers to historic satellite imagery and quantifying accuracy was not possible in the given time frame. As the challenges in achieving the original aims of this research project were realised, lines of enquiry were altered to more thoroughly investigate novel areas of the workflow where no sufficiently detailed literature exists.

This thesis describes the data preparation and classification methods developed as a foundation on which research into data comparability solutions and classification optimisation methods can be built. Concisely, the three most important outcomes of this work were:

1. Development of a method for preparing classifier training datasets compatible with both the source data (Landsat imagery) and the classifier (a Convolutional Neural Network based hybrid method implemented through Trimble eCognition) for achieving optimal classification accuracy.
2. Development of a classification framework that draws a compromise between the two aims of classifying land cover with high accuracy and classifying land cover with environmentally relevant detail.
3. Development of a functional machine learning classifier able to detect the desired land cover classes in Landsat imagery, including those that are not visible to the human observer.

Importantly, this thesis also describes the major barriers to further development of a method for producing an accurate time series of land cover change. The two most important problems encountered are:

1. Classifier models trained to detect land cover classes with modern Landsat imagery did not achieve any level of useful classification accuracy when applied to historic image datasets.
2. Accuracy assessment of classified maps produced by application of pre-trained classifier models to significantly older datasets proved to be incalculable through conventional methods.

In overview, this thesis should serve as an easily digestible resource to assist in future development of a software solution that can produce classified, time-series maps of native woody vegetation on New Zealand's sheep and beef farmland at a low cost. The intended purpose of this software is to assist Beef and Lamb New Zealand in achieving the goals set out in their Environment Strategy and Implementation Plan so that they can better support their farmers in acting as effective kaitiaki of the land and remaining compliant and self-regulating as environmental policy in the agriculture sector develops.

Acknowledgements

This thesis was only possible because of the support of the individuals and organisations that supported me financially, academically and socially during my studies.

Thanks to David Norton for mentoring and directing my interests and abilities towards a research project that both inspired and challenged me. The last two years have provided me with the knowledge and confidence to make a living pursuing my fascination with the biogeography of New Zealand.

Bradley Case and Graham Hinchliffe from Auckland University of Technology for their GIS and remote sensing support. You guys deserve a holiday.

Vega (Cong) Xu for her local GIS and remote sensing support. Thanks for listening to me while I verbally confused myself and clarified my own problems whenever I came to you for help.

Justin Morgenroth for his compelling teaching style and enthusiasm for the power of GIS. Taking FORE642 made learning GIS close to effortless.

Bruce Manley for all the background support and for somehow, against all odds, managing to teach me some principles of forest economics.

Jeanette Allen the solver of all problems in the School of Forestry. We all have Jeanette to thank for our success, without her institutional knowledge and indomitable attitude, life as a postgraduate student would be so much harder.

Beef and Lamb New Zealand and my contacts Tom Orchiston and Suzanne Keeling for their massive financial support and providing notes on my writing.

The WIDE Trust and my contact Sue Patterson for their generous support in paying my course fees.

Mum and Dad, for raising me to be curious and hopeful. Abe and Astra for listening to me rant and rave. It's all part of the process and I couldn't imagine a better way to grow up.

Sarah. Buddy we will make it.

Garry and Shar Parker for welcoming me into their home so warmly during CoViD and crunch time. Things couldn't have worked out better.

Shailer and Candice for letting me rent a wee slice of Haven. Every day in the valley is a good one.

All my friends and whanau who have been keeping me from unintentional social isolation and helping me fix my formatting, fix my programming, fix my car, build computers and fish and hunt for kai. I could have never come so far without this network.

Table of Contents

List of Figures

List of Tables

Definitions and Abbreviations

Historic Imagery – Satellite or aerial imagery with an acquisition date earlier than the most recent (primary) satellite image from which the classifier was trained and to which a pre-trained classifier may be applied.

Patch/Fragment - Any discrete area of land covered in vegetation that can be represented as a polygon in GIS software. The smallest size of a patch in this research was equal to one Landsat pixel (30x30 metres).

Remnant - Specifically refers to a patch/fragment that contains elements of the original (pre-clearance) old growth native forest. Remnant forest can include forest which was incompletely cleared or selectively logged but retains elements of old growth forest. A key defining feature is that native forest has always been present on the site.

Old growth - Specifically any native forest type or component (element) of that which is likely to have been continuously present on the site since before the arrival of humans to New Zealand (approx. 1280 CE).

Regenerating or successional forest – Native forest that has established on an area of land that was, at one time since the arrival of Europeans to New Zealand, cleared of the original native land cover.

AOI – Area of Interest.

AUT – Auckland University of Technology.

DN – Digital Number. The radiation data values recorded by the sensor after initial on-board processing. Representative of the reflectance spectra emitted by the land cover after travelling through the Earth's atmosphere.

ESRI – Environmental Systems Research Institute. Developers of ArcGIS software.

ETM+ - Enhanced Thematic Mapper +. The sensor carried on the Landsat 7 mission.

ETS – Emissions Trading Scheme.

GIS – Geographic Information System

KIA – Kappa Index of Agreement

LCDB – Land Cover Data Base (Landcare Research).

LINZ – Land Information New Zealand.

LENZ – Land Environments of New Zealand. (Ministry for the Environment).

LEP – Land and Environment Plan (Beef and Lamb New Zealand).

MfE – Ministry for the Environment.

MSS – Multi Spectral Scanner. The sensor carried on Landsat missions 1-5.

NIR – Near Infrared.

NES - National Environmental Standards.

NPS - National Policy Statements.

NWV - Native Woody Vegetation. In this report is used to refer to both remnant and regenerating native forest and shrubland.

OBIA – Object Based Image Analysis.

OLI/TIRS – Operational Land Imager/Thermal Infrared Sensor. The sensor array carried on the Landsat 8 mission.

RGB – An abbreviation of Red/Green/Blue, this refers to standard, 3 band, colour imagery.

SR – Surface reflectance. The data values attained after atmospheric and topographic correction. Representative of the land cover's spectral profile as the light leaves the surface.

SWIR – Short Wave Infrared.

TM – Thematic Mapper. A sensor used on Landsat missions 4 and 5.

UC – University of Canterbury.

USGS – United States Geological Survey.

1. Rationale and Background

The original aim of this research was to produce a time series of classified maps through analysis of satellite data attained from the Landsat programme. The classified maps would describe quantitative, qualitative and spatial changes in Native Woody Vegetation (NWV) land covers that have occurred in three sheep and beef farming regions of New Zealand over the last twenty-five to thirty years.

1.1 Motivations for this research

Part of the impetus for this research is the increasing force of environmental regulations being applied to businesses and landowners in New Zealand. Changes to proposed regulations such as the Emissions Trading Scheme (ETS), National Environmental Standards (NES) and National Policy Statements (NPS) are likely to impact many landowners over the next few years as New Zealand looks to address issues of land use sustainability and environmental degradation on local to national scales. Geographic analysis of land cover composition and its change over the last thirty years is an important tool for measuring the quantitative and qualitative changes associated with these environmental issues and thus providing a baseline for informed future policy development. Changes in shape, size and composition of patches of vegetative land cover can allow land managers to assess an area's environmental qualities at a glance and digitally demonstrate the effects of long-term management strategies in detail.

As well as the pressure of a changing regulatory environment, consumer pressure for information is an increasing force that is driving change in the way that meat and fibre products are produced and sold. Consumers are becoming more discerning about the chain of custody of the products they buy, and the market for products sold with certification ensuring a particular level of environmental or ethical responsibility is increasing. As New Zealand has identified areas of environmental policy that require improvement, it has simultaneously become more important to people to be able to track the history of their consumer products. Development of a more robust system for quantifying the results of environmental management actions will benefit farmers by allowing them to remain ahead of regulation and provide a more desirable product to the consumer.

In May 2018 Beef and Lamb New Zealand formalised their motivation to improve the environmental impacts of sheep and beef farming by publishing their "Environmental Strategy and Implementation Plan 2018-22". The products of this research have the potential to contribute to all five of the objectives in their vision statement (Beef and Lamb

New Zealand, 2018) through development of tools for measuring recent land cover change in a way that is relevant to the quality of freshwater, biodiversity and carbon balance. Achieving these objectives will put the sheep and beef sector in good stead to remain compliant and self-regulating as environmental policy develops over the medium term.

Specifically, development of these measurement tools and the subsequent production of detailed land cover maps will directly improve farmers' ability to:

1. Optimise the natural resources of their farms to profitably produce high quality food and fibre.
2. Create a sound Land and Environment Plan (LEP) that is easy to keep current as the template is updated.
3. Be transparent about their environmental challenges and the ways in which they are addressed.
4. Ensure their farming landscapes are biologically diverse, freshwater quality is protected, soils are healthy, and the sector has a carbon footprint that is sustainable in the long term.
5. Demonstrate their contribution to environmental goals.

These same tools will contribute to advancement in ecology and remote sensing as there is an identified lack of ability to assess the quality and spatial arrangement of patches of NWV, especially in low altitude land environments and within the matrix of exotic grassland and forest found on farms (Norton & Pannell, 2018). One of the best lines of investigation for improving methods of assessment of these attributes is through analysis of the Landsat catalogue, due to its long-term of operation and free availability of data. There is little precedent in the literature for measuring long term land cover change from Landsat data over relatively small areas. Studies such as Phiri et al. (2019) and Vittek et al. (2014) carried out time series analysis of land cover using Landsat data, but the studied areas were much larger. The extraction of detailed data products from this spatially coarse but spectrally fine resolution data would be both a novel step in remote sensing research and a powerful tool for addressing the modern day challenges of managing land sustainably in New Zealand.

Specific Changes to Land Management Policy and Regulation

Significant development of environmental policy concerning land management has occurred over recent years and implementation of new regulations to align with this will occur over the next 5 to 10 years as Councils' district and regional plans are updated. These changes are likely to have considerable impact on all businesses that utilise large areas of land and sheep and beef farms will undoubtedly be affected.

The first policy change of concern aims to reduce greenhouse gas emissions from agriculture, most likely through inclusion of the agricultural sector into the ETS. The Action on Agricultural Emissions discussion document (Ministry for the Environment, 2019) details proposals for pricing and penalising farmers for carbon emissions from livestock and fertiliser. Alongside this there is an indication that farmers will receive pricing and crediting for carbon sequestration on top of a “free” baseline of permissible emissions. This regulation proposal is still in the early stages and pricing is expected to be set no earlier than 2025. Compulsory emissions reporting however is expected to be implemented by 2024. As this is a discussion document there is still some uncertainty around the expected time frames and there is a contingency option in which pricing would instead be implemented at the processor, rather than the farm level by 2025. The Interim Climate Change Committee (ICCC) is a ministerial advisory group charged with managing the goal of reducing carbon emissions from agriculture. Discussions between this group and the agricultural sector are still ongoing, but it seems likely that the goals and timeframes provided will be a reliable, although approximate, guide to future developments in New Zealand’s greenhouse gas emission regulations.

The introduction of a new policy in the form of a NPS on Indigenous Biodiversity will also compel change in the environmental management practices of all land managers, including sheep and beef farmers. The proposed policies and objectives of the Draft National Policy Statement (Ministry for the Environment, 2019) are numerous but generally concern maintenance, restoration and management of indigenous biodiversity on private land. The policies and objectives also account for the role of the people (“landowners, communities and tangata whenua” – Ministry for the Environment, 2019, p. 15) as kaitiaki of the land and the intention of this policy is to support the people’s ability to manage and benefit from maintenance and restoration of biodiversity. These proposed policies will likely designate most areas of NWV on sheep and beef farms as environmentally (and legally) significant, restricting certain land management activities (e.g. land clearance) through updated district plan regulations. The timeframe for monitoring of the effect of implementation of this policy is not specified in absolute terms but is intended to occur within ten years of implementation of the finalised NPS.

Amendments to the National Policy Statement for Freshwater Management 2014 (NPSFM) will also impact sheep and beef farmers. Although originally published in the New Zealand Gazette in 2014 and amended in 2017, finalised water quality targets were made available to the public by all regional councils by the end of 2018 and management programmes to meet these targets are expected to be implemented by 2025 or 2030 in special circumstances. An independent review of the implementation and effectiveness of the NPSFM is due to be completed by July 2020 at which point the Minister for the

Environment will consider further amendment to this policy (Ministry for the Environment, 2017).

Each of these policy changes and proposals rely on the development of tools to audit the carbon emissions, biodiversity values and water quality effects that result from the interaction of farming activities with the biotic and abiotic environments of terrestrial New Zealand. These tools will certainly contain a remote sensing and Geographical Information System (GIS) component. The research described in this thesis is intended to provide guidance for further investigation on how farmers can enhance their ability to understand, communicate and improve specific environmental effects of their business activities through better understanding the history of NWV on their farms without impacting their ability to profitably produce meat and fibre.

1.2 Background

Deforestation in New Zealand

Anthropogenic deforestation of New Zealand is a comprehensively studied topic. It is estimated that before the arrival of humans in approximately 1280 CE, 80-85% of New Zealand's land area was covered with native forests. Since then, waves of human migration from Polynesia and Europe drove the clearance of the vast majority of this forest and today, native forest is estimated to cover just 23-30% of the country's land area (Ministry for the Environment, 2018; Norton and Pannell, 2018).

Forest clearance was, however, not an unbiased process and the forests which covered the low altitude plains and hill country, especially in the dryer east of the North and South Islands, were removed first and most completely. Evidence suggests that much of the deforestation was caused by accidental ignition by early Polynesian settlers and later the deliberate burning and felling by European settlers (McGlone, 1983; Perry et al., 2012). In some places this clearance was almost complete, and the majority of the remaining forested land area is comprised of scattered fragments. Due to their rarity and remoteness, the composition and environmental function of these fragments is often poorly understood (Norton et al., 2020).

In the regions where native woody vegetation is least abundant overall, fragments of lowland forest on sheep and beef farmland are disproportionately important reservoirs of native biodiversity when compared to public conservation land (Norton and Pannell, 2018). Of the eight Land Environments of New Zealand (LENZ) classes noted in Norton and Pannell (2018) as having less than 10% of their land area covered with native woody vegetation, six of them have more native woody vegetation on sheep and beef farmland than on public conservation land. Most of these classes have at least twice as much forest on sheep and beef farms as on public conservation land and the forest that occurs on sheep and beef farms represents a stronghold of the forest types that have been most impacted by human clearance in New Zealand. The forest communities endemic to these areas are at the highest risk of further degradation and loss of biodiversity, as the legal and practical protections that can be afforded to them are minimal and their diminished size is an innate vulnerability (Ewers et al. 2006; Deconchat et al. 2009; Monks et al. 2019).

Forest Remnants

The presence of old growth or remnant forest is an important attribute when it comes to determining the biodiversity value of a patch of native woody vegetation (Forbes et al. 2020). The extent of old growth presence in a patch can have huge implications for the ability of the patch to support a mature and diverse forest community including plants and animals, or act as a propagule source for a rich selection of plant and animal species. At present, old growth forest cannot be reliably discriminated from regenerating, successional forests with the remote sensing methods used to produce the Land Cover Database (LCDB) land classes. Only ground based surveys and historical records are able to produce any kind of reliable estimate of the extent of old growth forests.

Forest patches on sheep and beef farms can be generally described as existing in a matrix of exotic grassland which is of low value as habitat for many native species. A consequence of this is that the biodiversity value of a patch of forest is not only defined by size and forest type, but also by continuity and arrangement. The continuity of one forest patch with others of similar ecology can strongly influence how useful its natural resources are to native flora and fauna. Features indicative of native woody vegetation patch continuity are visible in the classified maps produced in this study (pp. 62-63,70-71,80-81) and the relevance of these to land owners in the face of changing environmental policy are discussed on pages 1-4 (cf. Franklin & Lindenmayer, 2009).

Land Cover Change in the Present Day

Today, exotic grassland is the most extensive land cover in New Zealand. The LCDB estimates that total exotic grassland covers 10.7 million hectares, equating to almost 40% of New Zealand's total land area (Ministry for the Environment, 2018). Estimates produced through an approach that combined LCDB data with the LENZ classifications and a number of other datasets (Norton & Pannell, 2018) show that sheep and beef farmland covers 40% of New Zealand's land area, while dairy farmland represents an additional 10%. Combined, this large area of farmland holds a significant amount of NWV, mainly in gullies, riparian strips and remote areas.

In an analysis of change to New Zealand's indigenous land cover and its level of legal protection, Cieraad et al. (2015) showed that the land area classified as being covered with native vegetation (including grassland and shrubland) had consistently decreased over the period 2002 to 2012. Furthermore, they showed that warm, dry and flat land environments (usually in low lying, eastern parts of the country) where the least native land cover existed were most likely to have lost indigenous land cover than those areas with harsher (colder, wetter, steeper) environments.

Analyses presented by the Ministry for the Environment in their most recent summary report (Our Land, MfE, 2018) indicate that in the period 1996-2012 both indigenous forest and exotic grassland cover have decreased slightly overall. A closer examination shows that exotic grassland contracted considerably in the periods 1996-2001 and 2001-2008, and was mainly replaced by exotic forest and cropping/horticulture. But in the 2008-2012 period exotic grassland showed the largest increase of any land cover class while total change of land use occurred at the lowest volume recorded. Overall, these small time-scale analyses show that although the rate of land use change appears to have slowed in recent years, New Zealand's balance of land cover is still a highly dynamic system that generally favours protection of certain land environments over others (Cieraad et al., 2015).

Furthermore, data from the most recent Mainland New Zealand LCDB estimates (v.5.0) were released early in 2020 (Landcare Research New Zealand Ltd, 2020). Land cover area estimates in this database are nominally dated at 2018 and generally support the observed trend of loss of indigenous land cover since 2012. While one class of NWV (Broadleaved Indigenous) did show an increase, all others had reduced in area. Detailed analyses of these data have not yet been released and the implications of these high level changes to patches of NWV and their environmentally relevant spatial attributes are not yet clear. An analysis of these changes as they relate to New Zealand Sheep and Beef farmland would be valuable to this topic of research and could support future development of time series maps of historical land cover change.

Social and Political Drivers of Change in Land Cover

Although the reasons for and extent of historic deforestation resulting from human settlement are well described in the literature, there are few descriptions of the changes that occurred in more recent decades. This section will explore some of the potential social and political changes of the last 80 years that are likely to have contributed to identifiable changes in the satellite and aerial photography record.

After the conclusion of the Second World War, the government allocated blocks of land to returning servicemen. This event caused a spike in the conversion of native forest to farmland mostly in marginal areas where crown land could be partitioned off easily (Ewers et al., 2006; Taylor et al., 1997).

The Korean War in the 1950s provided a vast market for New Zealand's wool and again the conversion of areas of native forest to pasture for sheep farming increased. While the demand for wool was the initiating factor for this spike in land cover conversion, the 1950s marked the start of the last phase in the history of New Zealand's native timber industry (Taylor et al., 1997).

In 1985, the New Zealand government implemented agricultural policy reform as a part of the response to the sustained and increasing budget deficits that had occurred over the preceding decade. Long standing agricultural subsidies were removed and animal product markets were deregulated in a process that wiped out many facets of financial support that were built in to the profitability models of New Zealand's farmers (Gouin, 2006). The election of a Labour government in 1984 likely helped cement this policy change, as much of Labour's voting base came from the urban population and the voice of the rural community was of little consequence (Taylor et al., 1997). It was around this time that areas of sheep and beef farm land around the country began to experience increased regeneration of native woody vegetation, partially as a consequence of this policy shift (Norton et al., 2018).

Sheep and Beef Farming and Environmental Management

The beef and lamb sector has already demonstrated a willingness and ability to respond to shifting forces in land use and management. Since 1990, the total number of sheep farmed in New Zealand has dropped by 52% and total land area used has fallen by 28%. Over the same time period the sector's contribution to New Zealand's gross domestic product has doubled, an increase equating to around five billion dollars. Additionally greenhouse gas emissions and nitrate leaching per kilogram of saleable product have decreased by 40% and 21% respectively, an achievement that far exceeds the stated goals (Beef and Lamb New Zealand, 2018). This trend of environmentally conscious land management outcomes looks set to continue if the goals set out in Beef and Lamb New Zealand's recent Environmental Strategy and Implementation Plan are followed to fruition.

Alongside the responsibility to manage native woody vegetation fragments, sheep and beef farmers are increasingly being asked to manage their soil and water in sustainable ways, specifically regarding nutrient leaching, hillside erosion and sediment discharge. While there are prima facie benefits of efficiency for farmers that manage these risk factors well, this is a clear space where creation or management of native vegetation fragments can produce multiple benefits, and encouragement should be given to any land manager looking to solve their agricultural problems through methods that contribute to biodiversity conservation in their catchment and greater region.

1.3 Thesis Aims and Outcomes

This research initially aimed to build upon the work of Norton and Pannell (2018) by producing a series of classified maps to describe changes in land cover on sheep and beef farms back through time. The original objectives stated in the thesis proposal aimed to use the Landsat catalogue and publicly available aerial imagery to achieve this by developing a method of automated land cover classification. This time series was intended to allow estimates of land cover change to be made from images captured prior to the development of the first iteration of the LCDB in 1996. The LCDB is a valuable resource for tracking land cover change, but it relies on contemporary ground truth data to produce accurate estimates of land cover status. The development of a tool to estimate the status of land cover from historical Landsat imagery in the absence of reliable ground truth data was the overall objective that this research aimed to contribute to.

Classified maps of modern Landsat imagery (2014-2016) were produced with moderate success, but the methods used to produce these did not achieve high enough classification accuracy when applied to historic Landsat imagery (1989-1990) to allow a change detection time series to be produced. The aims of this thesis were revised to identify and address the specific issues that caused decreased classification accuracy in the modern Landsat imagery and prevented the successful transfer of trained classifier models to historic Landsat imagery. Figure 1 (p. 10) describes how the original objectives of this study were developed and refined into the objectives achieved in this research and the future objectives identified by this research.

The objectives achieved by this research are detailed in the methods and results sections and are the focus of this thesis. These objectives largely aimed to improve classification accuracy of maps produced from modern Landsat imagery as this was determined to be critical in producing any time series. The most important results of the work carried out in this thesis are:

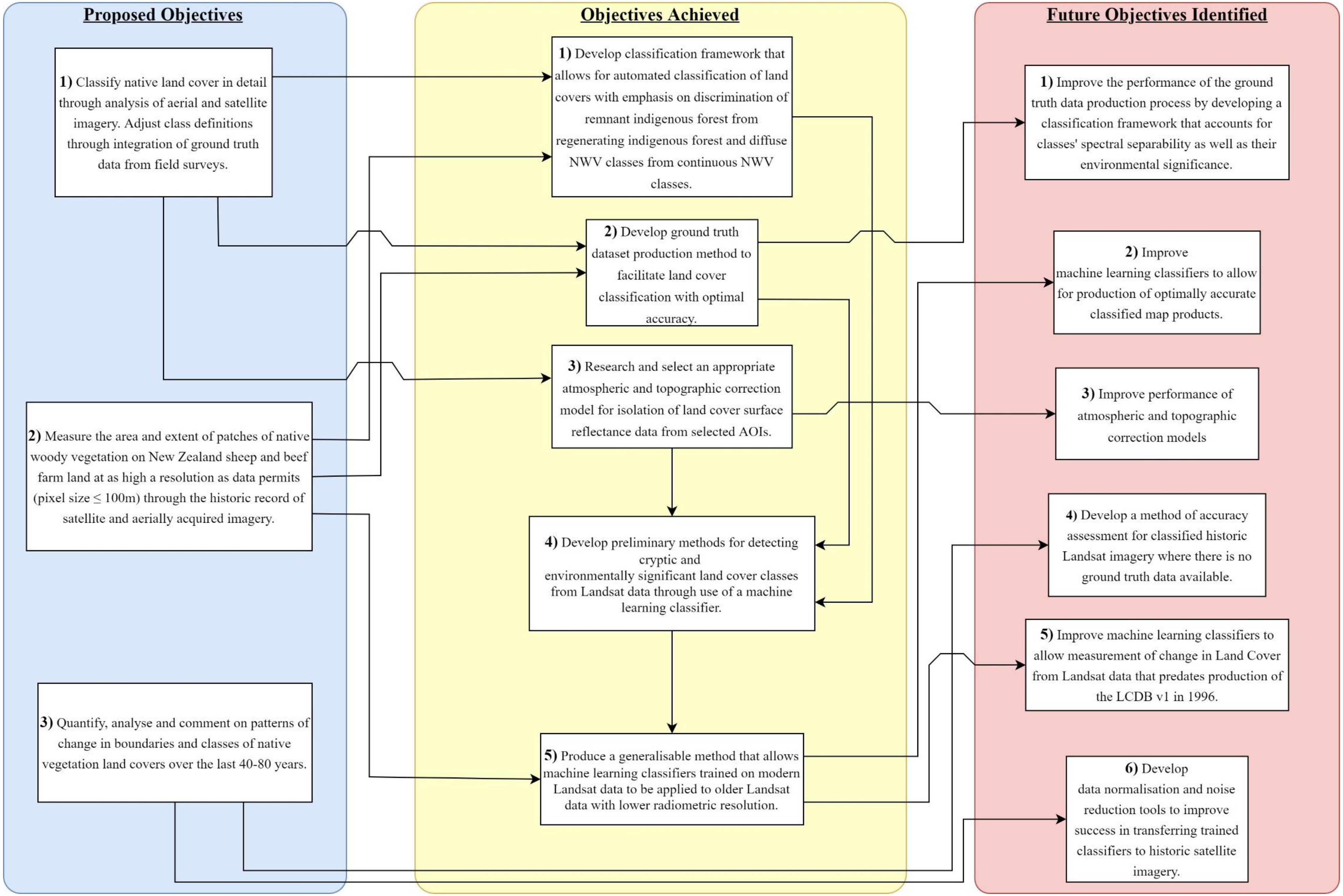
1. Development of a classification framework able to describe differences in density and ecological history of NWV while accounting for the classifier's ability to discriminate classes from the six most common bands of Landsat data.
2. Development of a standardised and effective method of producing ground truth data.
3. Development a functional automated classifier model and a statistical analysis of its success.

The future objectives identified are detailed in the discussion and conclusion sections and were unattainable in the time allowed for this thesis. These objectives are focused on developing data preparation and classification methods that will allow classifiers trained to recognise land cover classes in modern Landsat data to identify those same classes in historic Landsat imagery. The most important aims identified for future research are:

1. Development of a method for reducing noise in historic satellite imagery
2. Development of the best method for equitably normalising these data and the modern satellite imagery to a common bit depth.
3. Development of a method of assessing accuracy in classified maps of historical satellite imagery in the absence of a sufficiently detailed ground truth dataset.

The overall result achieved by the revised objectives of this research was the construction of a broad foundation from which deeper, more specific inquiries into the identified future objectives can be launched. The basic workflow structure and commentary on selected methods and unexploited opportunities should allow a widely applicable software solution to be more easily developed. Figure 1 on the next page describes the ways in which the original objectives put forward in the thesis proposal were changed and refined into the outcomes and objectives listed above.

Figure 1. Process Diagram of Research Objective Refinement



2. Data, Hardware and Software

2.1 Study Areas

Field surveys were carried out between January and May 2019 to attain a ground truth dataset of polygons labelled by land cover class. These data are referred to as the primary ground truth dataset in this thesis. The areas surveyed were chosen due to the large proportion of land in use for sheep and beef farming, their membership to some of the most common sheep and beef farm classes (Beef and Lamb New Zealand. 2018) and their accessibility to the university. The surveyor had previously undertaken vegetation surveys on farms in each of the selected study areas so was familiar with the distribution patterns of vegetation on the landscape and the species present.

Although the land in the selected AOIs is not strictly representative of sheep and beef farms across New Zealand, the compliment of native and exotic land covers visible in each are typical for the region and are a good platform for development of a generalisable method for automated classification of Landsat data.

Figure 2.Location of Study Sites in Context

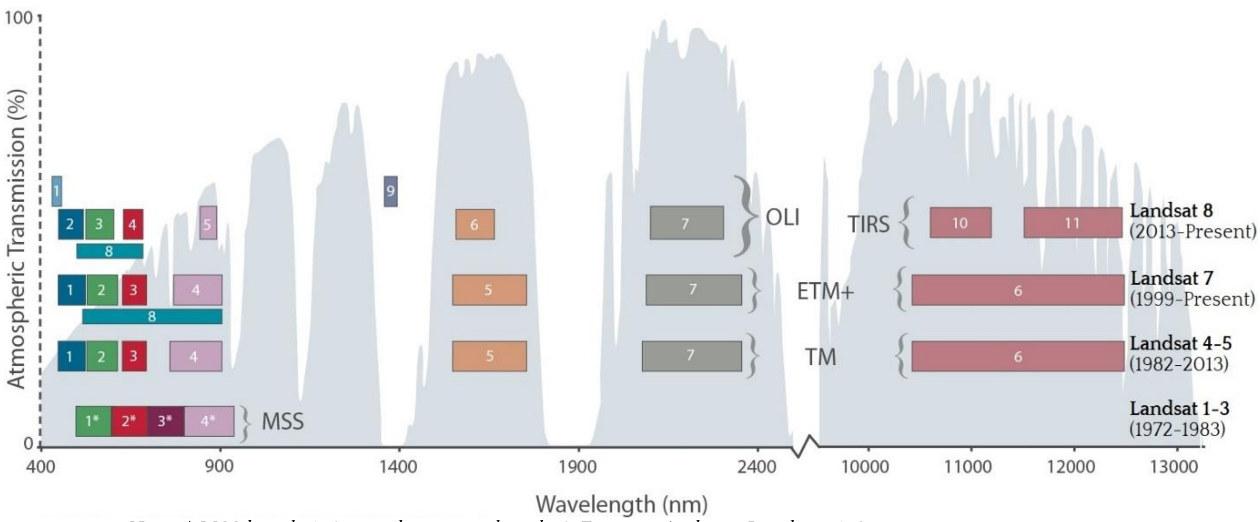
2.2 Data Description

Satellite Data

This research focused on analysis of data acquired by the Landsat series of Earth imaging satellites. Satellite data selection was restricted to the Landsat catalogue due to three important attributes: free availability, long period of operation and wide multi-band spectral range. These attributes were determined to be highly desirable in producing long term time series analysis of land cover through satellite data and the Landsat catalogue was the only data source that suited these parameters well.

From the catalogue of Landsat imagery a scene covering each Area of Interest (AOI) was chosen from two time periods: the Primary Landsat Scenes were aligned temporally with the most recent publicly available aerial imagery (2014-2016) and selected from the Landsat 8 catalogue; while the Secondary Landsat Scenes were acquired twenty four to twenty seven years prior and selected from the Landsat 5 catalogue. Although the image catalogues used have a period of sixteen days between capture of each image, most images from the desired time periods were unsuitable for use in this research due to cloud cover, digital artefacts, or missing data. Despite the vast amount of data available in the Landsat catalogue, random variation in local climatic conditions and digital sensor performance heavily restricted the availability of scenes suitable for land cover change analysis.

Figure 3. Comparison of the Range of Spectra Acquired by Each of the Four Generations of Landsat Sensor



*Note. * MSS bands 1-4 were known as bands 4-7 respectively on Landsats 1-3.*

Graphic created by Rocchio & Barsi. 2017. Public Domain. Modified with data from USGS (2020).

The sensors carried on the Landsat satellite platforms have evolved over time and each generation of sensor has acquired images at different data specifications. Figure 3 above compares the spectral ranges of each band captured with each of the four generations of Landsat sensors shown in order from oldest to newest up the y axis. The three most modern sensors: the Thematic Mapper (TM) carried on Landsats 4 and 5; the Enhanced Thematic Mapper (ETM+), carried on Landsat 7; and the Operational Land Imager/Thermal Infra-Red Scanner, (OLI/TIRS) carried on Landsat 8, are closely comparable in their spectral ranges. The oldest Landsat sensor however, the MSS, has a greatly reduced range of spectral capture (USGS. n.d.).

The spectral bands used in the land cover analysis were the six bands common to Landsat 4 and 5 (TM) and Landsat 8 (OLI/TIRS). These are designated as: Blue, Green, Red, Near Infrared, Short Wave Infrared 1 (SWIR 1) and Short Wave Infrared 2 (SWIR 2). These two sensor generations were launched thirty-eight years apart, and although these six bands are nominally common to both, there are some differences in the spectral ranges captured (see Figure 3 and Appendix C).

Key features of the selected satellite image datasets and land attributes of each study area (Area of Interest/AOI) are tabulated below.

Table 1. Key features of Selected Satellite Datasets

	Canterbury		Manawatū-Wanganui		Northland	
	Landsat 4 (TM)	**Landsat 8 (OLI/TIRS)**	**Landsat 4 (TM)**	**Landsat 8 (OLI/TIRS)**	**Landsat 4 (TM)**	**Landsat 8 (OLI/TIRS)**
Date of Acquisition	16/12/1990	26/12/2014	4/6/1989	10/9/2016	27/6/1989	16/8/2016
AOI area (ha)	77,897		88,392		64,672	
Level 1 LENZ Classes	B, E, J, N		C, D, F, H, P		A, D, G	
Farm Class	South Island Hill Country		North Island Hill Country		North Island Hill Country	

Note. Farm class from Beef and Lamb New Zealand 2018. LENZ classification from Ministry for the Environment 2019. LENZ class details can be found in Appendix E. AOI extents are shown in context in Figure 2.

Ground Truth Data

In this research, data used for ground truth was acquired from two sources: field surveys and aerial photography. These were combined to produce a polygon dataset, labelled by land cover class, for use in classifier training and mapping accuracy assessment. The primary ground truth data acquired through field surveys is described above in section 2.1. The aerial photography (referred to as secondary ground truth data) was used to enhance confidence in class nomination and boundary drawing when digitising and formatting the polygons drawn during field surveys. The dataset produced through combination of these two data sources is referred to as the combined ground truth dataset.

Secondary ground truth data was sourced from the Land Information New Zealand (LINZ) Dataservice, but was implemented in the GIS by streaming to ESRI ArcMap's "NZ Imagery" basemap data layer (LINZ. 2014-2015; LINZ. 2016-2017; LINZ. 2014-2016).

Table 2. Key Temporal and Spatial Attributes of Primary and Secondary Ground Truth Datasets

Region	Secondary Ground Truth Data		Primary Ground Truth Data
	Resolution	Capture Date Range	Date of Survey
Canterbury	0.3 metre	Summer 2014-2015	Nov. 2018- June 2019
Manawatū/Whanganui	0.3 metre	Summer 2016-2017	May 2019
Northland	0.4metre	Summers 2014-2016	May 2019

Note. (LINZ, 2019).

Ancillary Data

Topographic data in the form of contour lines was used in this research to produce a Digital Elevation Model (DEM). The contour line data is part of the NZTopo50 series and has a spatial resolution of twenty metres. This was retrieved from Land Information New Zealand (LINZ) Dataservice (LINZ. 2011).

Polygon data was produced institutionally. AOI extents were drawn during this research with guidance from Prof. David Norton. Sheep and Beef farm extent polygons were collated by Auckland University of Technology and supplied by Dr Bradley Case.

2.3 Computer hardware

Building automated classification models with a useful level of sophistication requires that the computer hardware used is well suited to each of the specific tasks in the workflows. Remote sensing and GIS software tends to be graphically demanding and the limitations of their use are often found to result from the hardware and software environments that they run within. Two different machines were used to run the machine learning classification algorithms that required high performance hardware. Specifications of the computer hardware used in this research are listed here.

Table 3. Key Computer Hardware Used

Component type	Specifications
Machine 1	
CPU	Intel Core i7 7700 @ 3.6GHz,4 Cores, 8 Logical Processors
GPU	NVIDIA GeForce GTX 1050; 2 GB GDDR5

RAM	32 GB
Storage	Various external HDD and SSD
Machine 2	
CPU	AMD Ryzen 7 3700X @ 3.6GHz, 8 Cores, 16 Logical Processors
GPU	NVIDIA GeForce RTX SUPER; 8 GB GDDR6
RAM	16 GB
Storage	Various external HDD and SSD

2.4 Software Selection

Software selection was a significant part of this research as several of the workflows required specialised software to carry out specific tasks. Software for image processing, ancillary data production, classification and analysis was chosen early in the research period, while software for data normalisation or correction took longer to choose. Software selections were finalised through trial and error testing for suitability to a task, but general criteria for initial exploration included:

- Ability to use a wide range of data types
- Specialisation to task
- Stability and user friendliness
- Compatibility with available hardware
- Licencing status at the University of Canterbury.

Final software choices are specified below.

Table 4. Notable software used

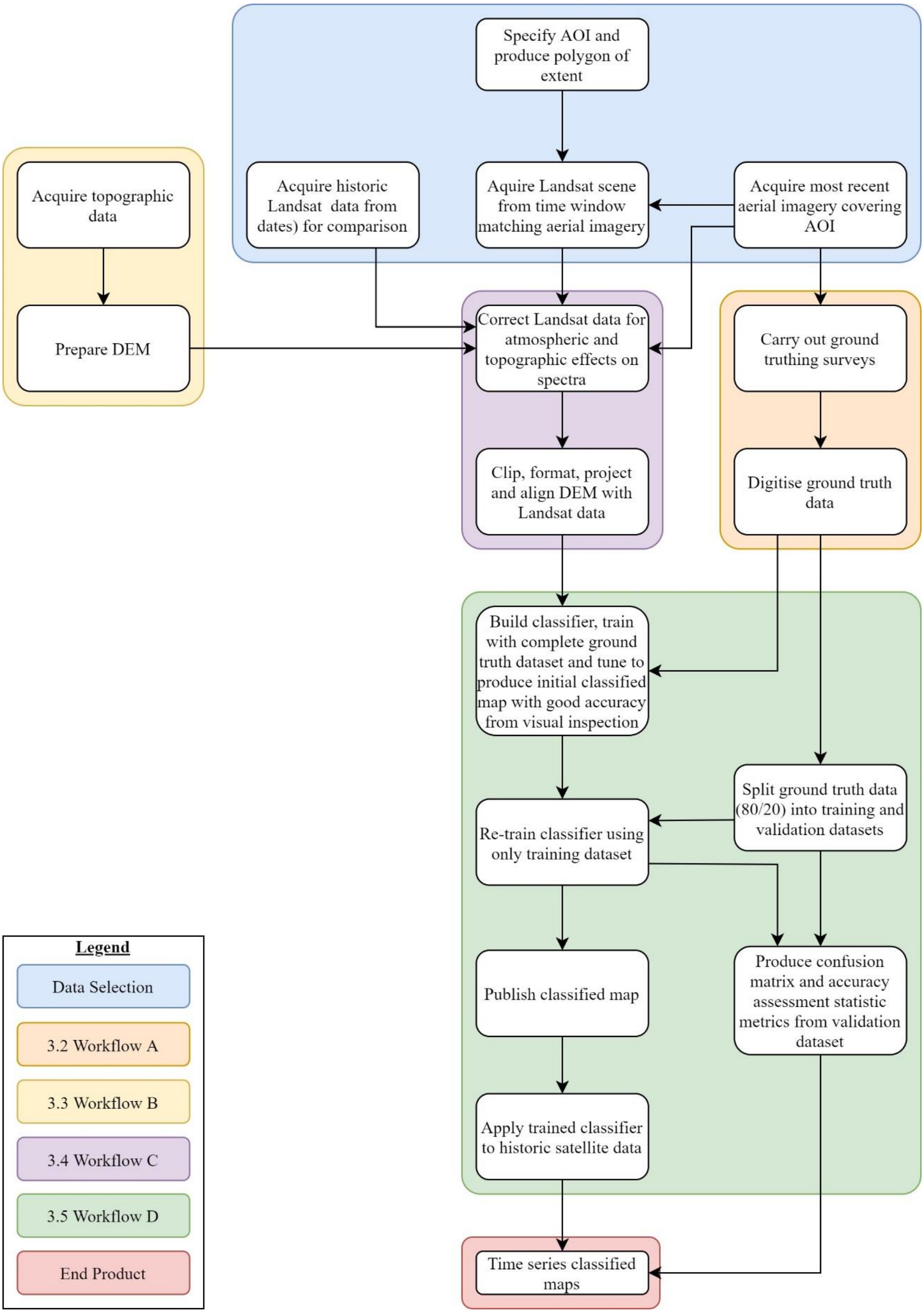
Name	Publisher	Use	Licence
ArcGIS Desktop 10.6	Esri Inc.	General geoprocessing and data management	Educational Licence (UC)
Collector for ArcGIS	Esri Inc.	Ground truth data collection	Educational Licence (AUT)
eCognition Developer	Trimble Inc.	Image classification	Educational License (UC)
ATCOR2/3	ReSe Applications GmbH	Atmospheric and topographic correction	Educational License (UC)
MODTRAN 5	Spectral Sciences Inc.	Atmospheric and topographic correction	Implemented within ATCOR2/3

3. Method

The methods presented here are a summary of the key actions taken in the final version of each workflow followed by descriptive paragraphs that provide details of each of the steps. Workflows were produced through an iterative testing process that required the researcher to test the effects of change in different functional components and alter the approach of the workflow in response to the attained results. As such, this methods section is not a precise record of the work carried out in pursuit of the project's aims, but a recommended workflow that is trimmed down to the most successful components (details of some important components of work that were excluded from the final workflows can be found in the discussion section p.94).

3.1 Workflow Overview

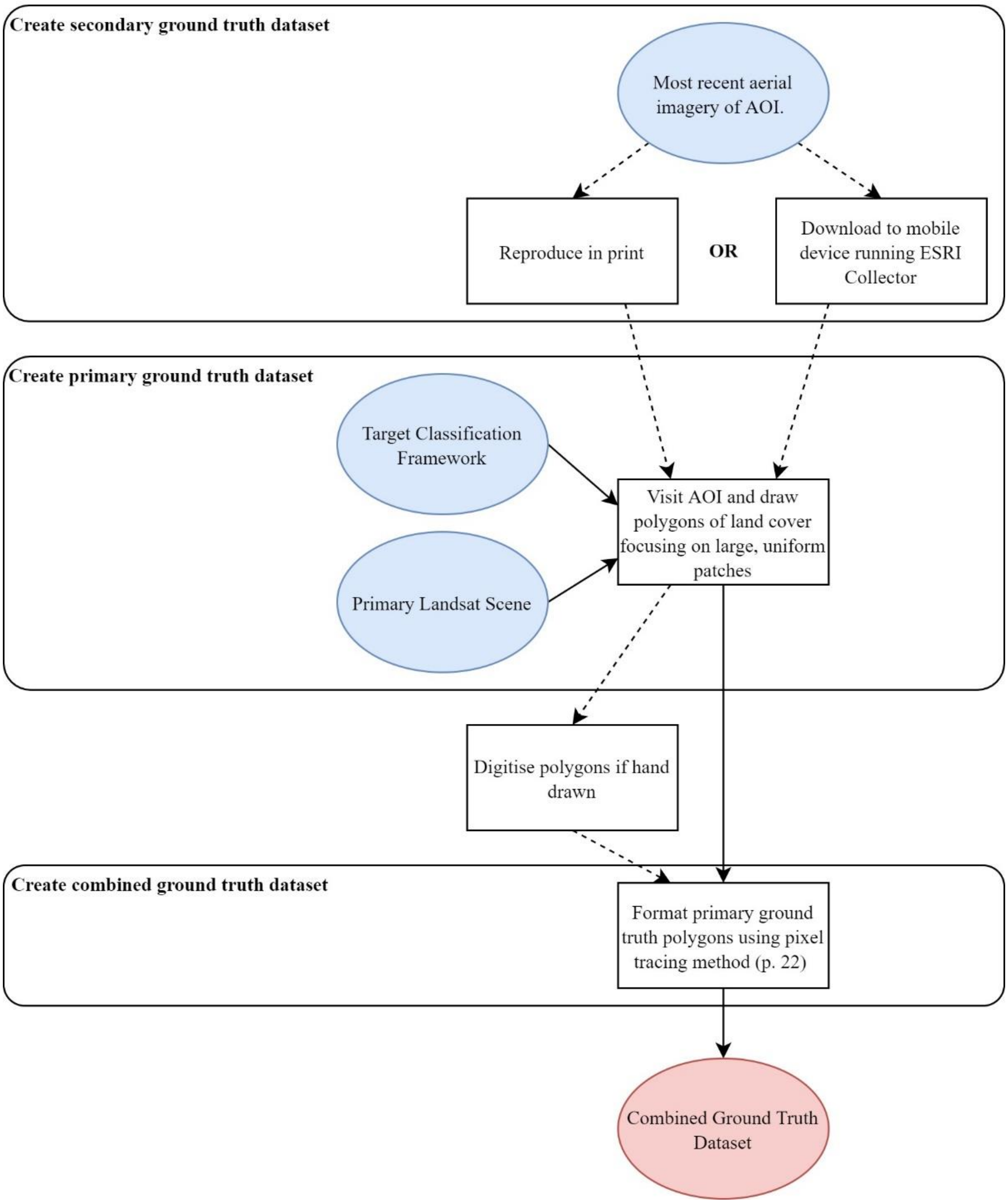
Figure 4. Summary of All Workflows and Their Relationships



3.2 Workflow A: Ground Truth Data Collection

This research used several sources to build the ground truth dataset used for training and validating a land cover prediction model (the classifier). The data selection and integration methods used to produce the ground truth dataset were designed specifically for use in this research project with applicability to the satellite datasets and classification software in mind. The ground truth dataset used: observations from physical site visits as the primary source of data; recent aerial orthophotography as a secondary source of data; and the target classification framework (Appendix A), and satellite data contemporary to the aerial orthophotography (Primary Landsat Scene) as supporting data. The steps taken to combine these data to create a dataset of high quality and applicability to this project are described in the flowchart and paragraphs below.

Figure 5. Ground Truth Data Handling Workflow



Survey Methods

Initial ground truth data was collected through traditional surveying methods, the researcher visited the AOIs and defined boundaries of target land cover classes by drawing polygons over the most recent and high resolution aerial orthophotography available. All surveying was carried out from the public roadside using binoculars to carefully assess the more distant patches of land cover. The fusion of the primary and secondary ground truth datasets was carried out coarsely in the field while the finer detail boundary drawing, class nomination and error reduction processes were carried out from the office later.

Most of the primary data collection was achieved with pen and paper but a significant portion of the later polygons were drawn using a mobile application: Collector for ArcGIS (henceforth “Collector”). Collector was implemented through use of templates provided by Graham Hinchliffe of Auckland University of Technology (AUT) and run on a Samsung Galaxy S8. For best results, a large mobile device such as a tablet should be used in combination with a stylus to allow for accurate vertex placement. This technology has the potential to reduce the amount of time taken in the digitisation and formatting phase significantly.

These surveys had two main purposes: to validate the land cover class of patches visible in the secondary ground truth data and to train the researcher to more accurately identify the class of patches of land cover through examination of secondary data only. It was important that the researcher was able to validate samples of the land cover classes in person as some were impossible to visually discriminate through examination of the imagery alone, regardless of the level of experience in image interpretation, these are referred to in this report as cryptic classes. Although an attempt was made to collect data in a strict systematic fashion, boundary drawing relies heavily on the researcher’s discretion for deciding whether or not to include a region in the primary ground truth dataset. Regions of uncertain, mixed or novel land cover class were recorded, but annotated with a description in case they were found to be useful later. These non-target class polygons were systematically filtered out in the digitisation phase as the classification framework and the methods of assessing sample quality were developed.

Large, uniform areas of each target land cover class were sought out and boundaries were drawn with a bias towards matching the patch edges visible on the orthophotography over the patch edges visible from the ground. Uniform areas were defined as any patch of land cover that fit the description of a single target class definition (see target class framework, Appendix A). Uniform land cover should not be confused with continuous land cover which is a density modifier of many class definitions (see target and actual class definitions, Appendices A and B).

Many patches of land cover that did not fit any of the target class definitions, or were of mixed class, were recorded during the survey process but excluded from the combined ground truth dataset. The only mixed class patches included were the diffuse woody vegetation classes that were interspersed with pasture and bare ground. Any diffuse patch of forest that was interspersed with a significant amount of other forest or shrubland classes was considered invalid for inclusion to the combined ground truth dataset due to the risk of introducing too much variation into the spectral data pool and reducing the ability of the classifier to define effective class membership thresholds.

Areas that had observable land cover change between the dates of primary and secondary ground truth data production were excluded from the combined ground truth dataset. Satellite imagery was chosen to match the land cover status on the date that the secondary ground truth data was captured, not the day of the primary ground truth surveys (dates of ground truth data production can be found in Table 1, p. 14). These areas were usually affected by an event that occurred over a large area in a short time period such as a plantation forest harvest, a fire, or cultivation of agricultural fields. These areas were considered invalid for inclusion into the combined ground truth dataset as there was not enough evidence to infer what land cover class existed on these areas at the time of satellite data acquisition from secondary ground truth data alone.

Satellite Data Selection

Choosing the base data for analysis is a critical step and any insufficiencies produced in this early stage of the mapping process will be propagated through the filtering and analysis phases of the project where they will produce errors, in some cases this may even have a dilatory effect and become an even bigger problem than anticipated.

Landsat scenes are available for download by the public from several online data portals. In this study all satellite imagery was downloaded from Earth Explorer, a portal run by the United States Geological Survey, a branch of the United States Department of the Interior (USGS. n.d.). Earth Explorer provides intuitive search and filtering functions that allow quick and easy download of Landsat scenes provided the desired AOI, time period and sensor type can be specified. It also allows for a number of different data products to be downloaded, each data product category has different levels of quality and post processing. In this research all data was acquired from the Landsat Collection 1 Level 1 data set as this has the widest time period of available data and only basic pre-processing applied.

The satellite data selection process was simple enough to be described without requiring this section to be written into a sub-workflow. Once the dates of secondary ground truth data (aerial photography) and the borders of each AOI had been finalised, search queries were sent to the data portal to find scenes that matched as closely as possible to the secondary data's period of acquisition while also encompassing the AOI fully and being of high optical quality (low error and low cloud). Optical quality could be somewhat difficult to assess through the data portal so an excess of scenes were downloaded and re-assessed at full resolution. Each Landsat Collection 1 level-1 data product contains a quality assessment band which was helpful in assessing comparative quality of similar images, especially in enhancing the visibility of cryptic error sources, most importantly cirrus cloud cover. Where multiple high quality images were available within the secondary ground truth data's period of acquisition, bias was given to the earliest image acquired on the assumption that this would reduce the chance of the image containing areas of sudden land cover change (e.g. Plantation harvest, fire, pasture cultivation).

Digitisation and Formatting (Pixel-Tracing)

Digitisation and formatting of the combined ground truth data was a time consuming but critically important task as the training and verification data must be representative of the actual land cover for a trained model to have any reliable predictive power. This entire process was repeated with each revision of the classification framework to ensure polygon objects were representative of the class descriptions.

Practical polygon shape drawing and accurate designation of object classes in the ground truth data were key factors in improving mapping accuracy in the final product. This phase of the research aimed to combine primary and secondary ground truth data through two principle actions:

1. Polygons drawn in the ground truth surveys were assessed for quality, and their shapes altered where necessary to ensure that the data was imported accurately into eCognition.
2. Class definitions were altered to ensure that all polygons in the combined dataset fit the classification framework in use at the time (see Appendix A and Table 6 for full descriptions of target and final classification frameworks).

Polygons were examined and considered to be quality if the enclosed pixels covered an area of uniform class and their edges matched sufficiently well with the corrected satellite data's pixel edges. These criteria were determined to be the most important as they ensured that the classifier and accuracy assessment algorithms received minimal noise from within each class dataset and that the vector layer created upon importation of the combined ground truth shapefile to eCognition was maximally accurate to the boundaries drawn in the supplied dataset.

Initially the polygons drawn in primary data collection followed the shapes of the natural landforms and land covers visible during surveying and in the secondary ground truth data. These polygons were used to train early versions of the classifier, but the smooth lines drawn in the surveys did not accurately represent the sampling processes' smallest unit of analysis (the thirty by thirty metre Landsat pixel) so sampling was inaccurate and these classifiers had poor predictive power. Patch edges visible in the satellite image raster data are never smooth, so the following method (henceforth the pixel-tracing method) was developed to assist in rapid redrawing of polygon boundaries to better conform to the raster data.

The pixel-tracing method allowed for rapid comparison of a range of data to enhance the user's capability to draw boundaries around quality land cover class samples. A stack of different data layers were used to achieve this, they are listed here from highest to lowest in the ArcMap Table of Contents:

1. Combined ground truth polygons
2. Polyline "fishnet", spatially aligned to corrected satellite data pixel edges
3. Select spectral indices (those used in this research are detailed in Appendix: D, examples are shown in Figures 8, 9 and 10)
4. RGB composite of corrected satellite imagery (histogram stretched).
5. Aerial Imagery (Secondary ground truth data)

This layer structure was used to redraw polygon boundaries, the natural lines were snapped to straight edges by using the fishnet layer as a proxy for Landsat pixel boundaries and using the spectral indices to increase confidence in pixel inclusion, exclusion and reclassification decisions. The aerial imagery layer was relied on most heavily for estimating the class membership and percent of coverage of individual pixels, especially on patch edges or where complex spatial arrangements of different classes existed.

A region of combined ground truth data containing several different land cover and density classes is shown in Figure 6 (p. 25) overlaid on the secondary ground truth layer. Large, single polygons typical of land cover patches of continuous density can be seen overlaying the areas of Continuous Native Remnant and Pasture. The gradient between these two classes is represented by small and variably sized polygons covering patches of continuous regenerating native, diffuse regenerating native, diffuse native remnant, and diffuse broadleaved shrubland. These polygons were classified by individual examination of the contents of each pixel to estimate the type and percentage of woody vegetation cover. Pixels shown in Figure 6 with no overlay were considered too ambiguous to include in the combined ground truth dataset and were not assigned a class. Images of the same region are presented in Figures 7-10 (pp. 26-29) to demonstrate some of the visualisation methods used to display the satellite imagery that improved the researcher's confidence in assigning a pixel's class membership.

Figure 6. Pixel-Traced Combined Ground Truth Data Sample Overlaid on Aerial Imagery

Figure 7. RGB Satellite Data with Pixel-Traced Ground Truth Boundaries (0.5 min/max Percent Clip Stretched)

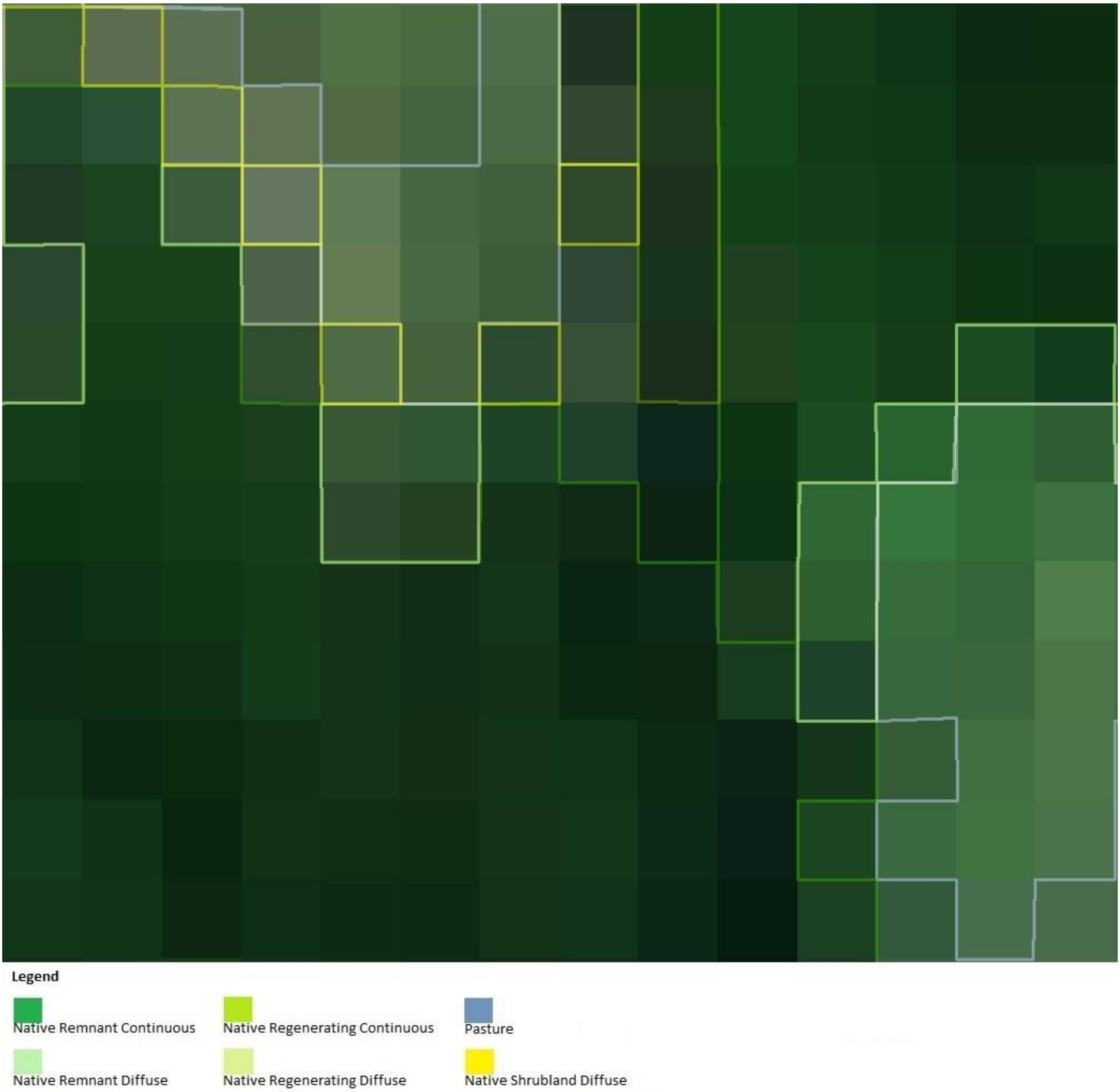


Figure 8. Wide Dynamic Range Vegetation Index (WDRVI) View of Satellite Data with Pixel-Traced Ground Truth Boundaries (Histogram Equalised Stretch)

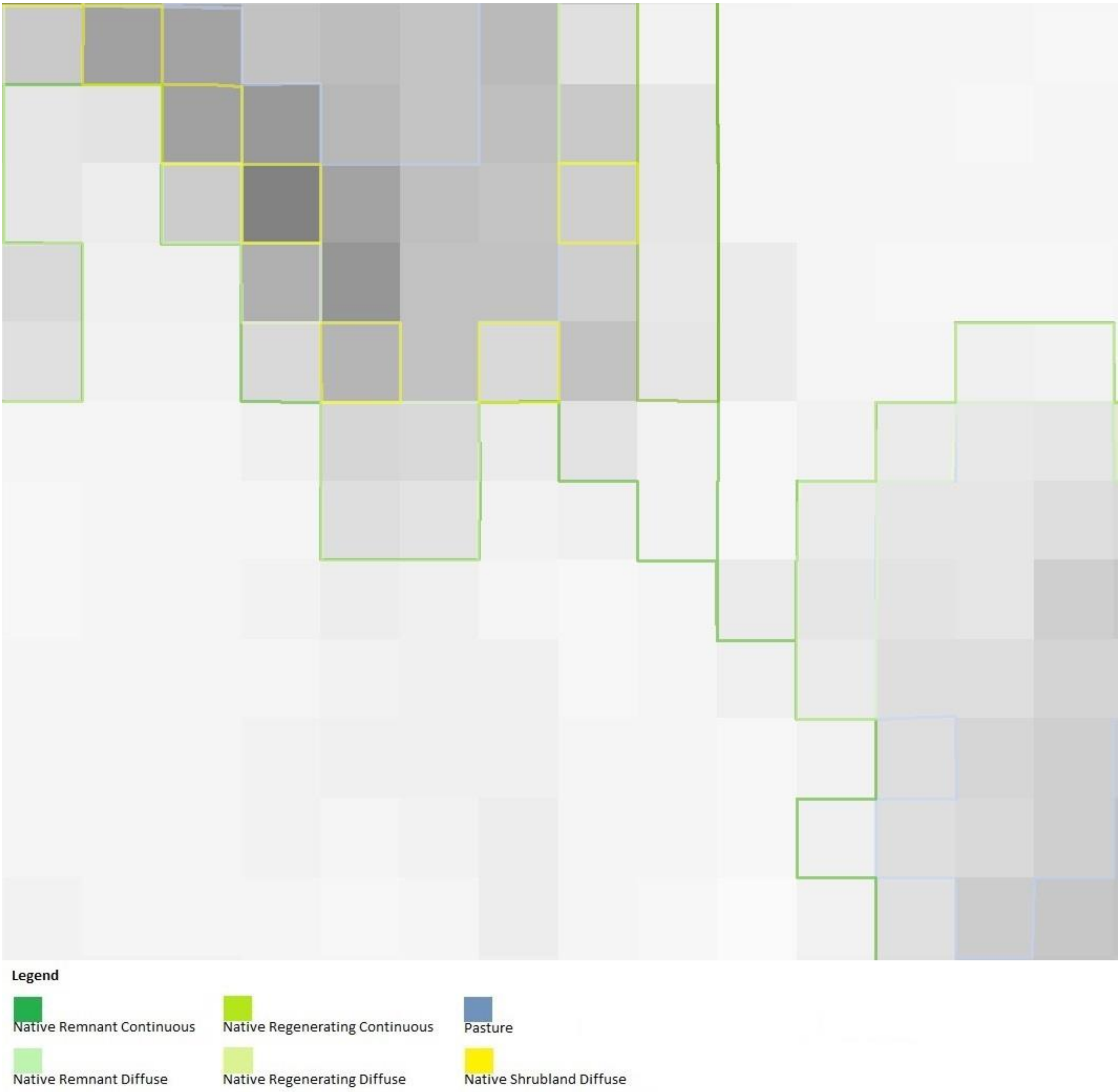


Figure 9. Soil Background Line (SBL) Index View of Satellite Data with Pixel-Traced Ground Truth Boundaries (Histogram Equalised Stretch)

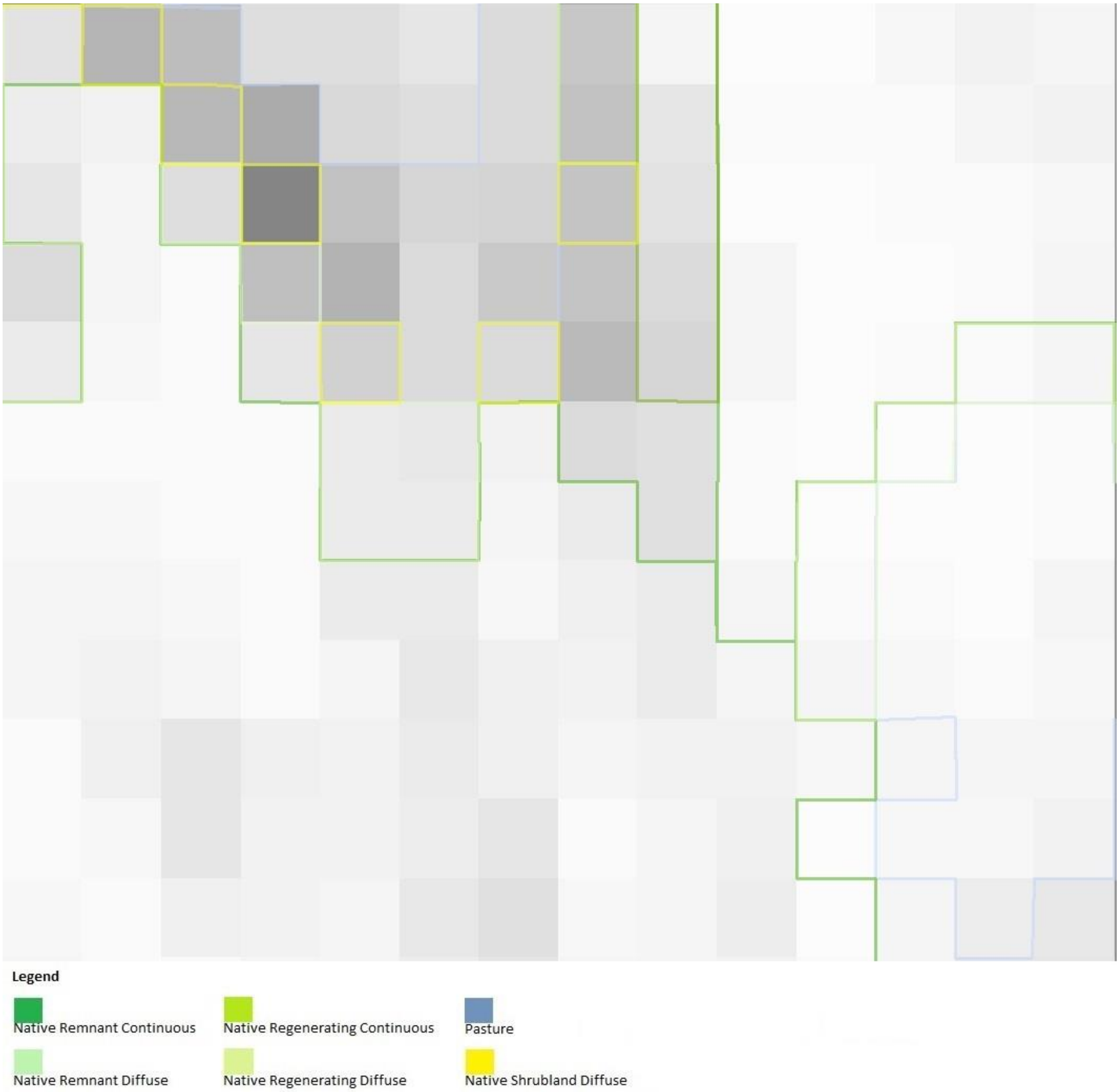
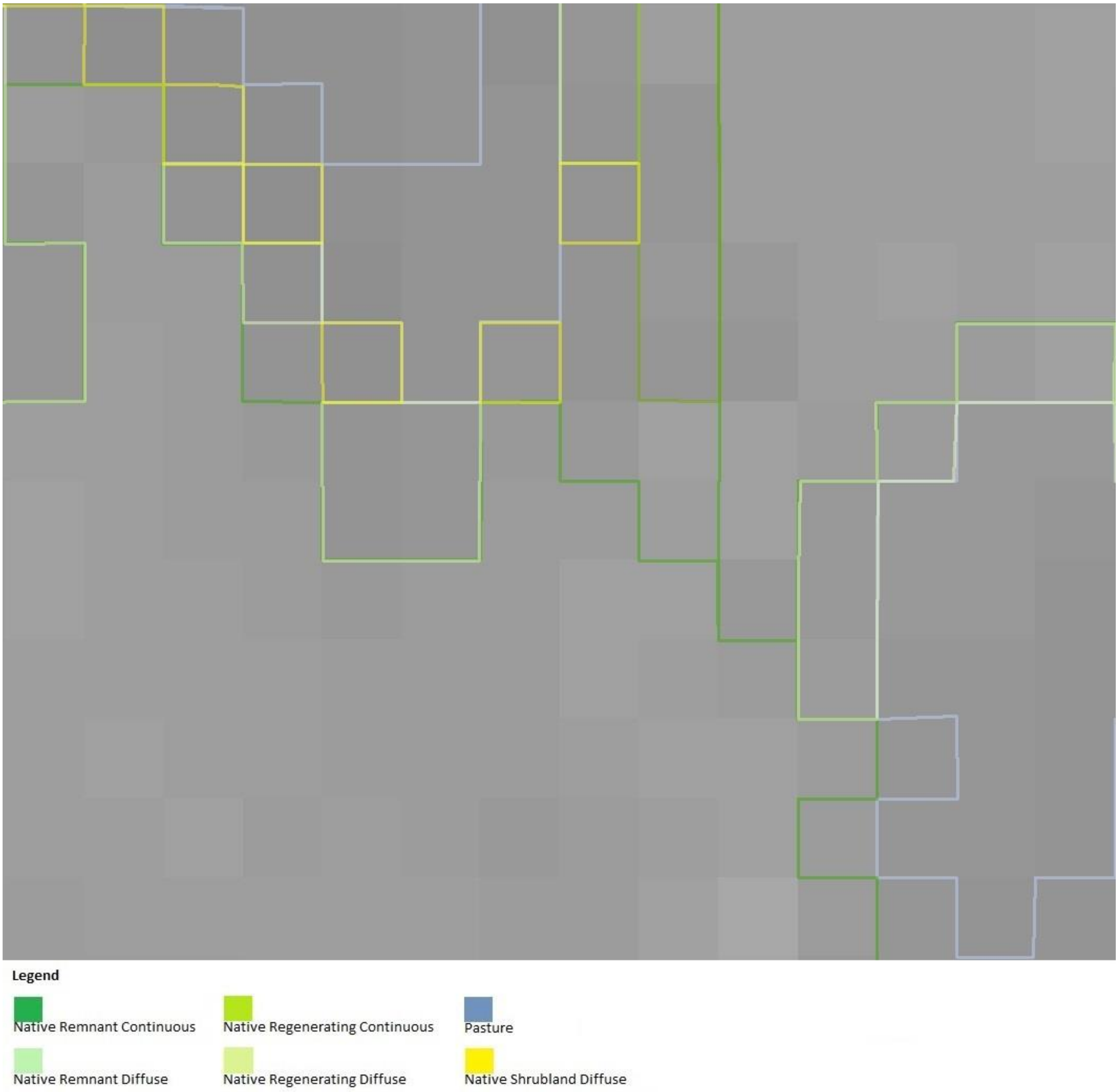


Figure 10. Visible Atmospherically Resistant Index (VARI) View of Satellite Data with Pixel-Traced Ground Truth Boundaries (4 Power Sigmoid Stretch)



Classification Framework

Initially the class definition structure followed the target classes (detailed in Appendix A) but this target classification framework was created prior to any detailed investigation into the discriminability of each class from the satellite data, and the classified maps produced from this class structure lacked accuracy. A series of modified classification frameworks were tested in an attempt to improve classification accuracy while retaining practical applicability of the data (significant permutations of these are detailed in Appendix B). This testing phase resulted in finding a point of compromise between the amount of detail in the class definitions (both in total number of classes and number of low density classes) and the accuracy of the classified maps created. Although changing the class definitions did improve classification accuracy, it also reduced the number of total classes ('sparse' density woody vegetation classes were removed) and therefore somewhat reduced the descriptive power of the output classified maps. The classified maps produced for this thesis report (Sections 4.6-4.8) are therefore necessarily a compromise between classification accuracy and class detail.

Sample Size

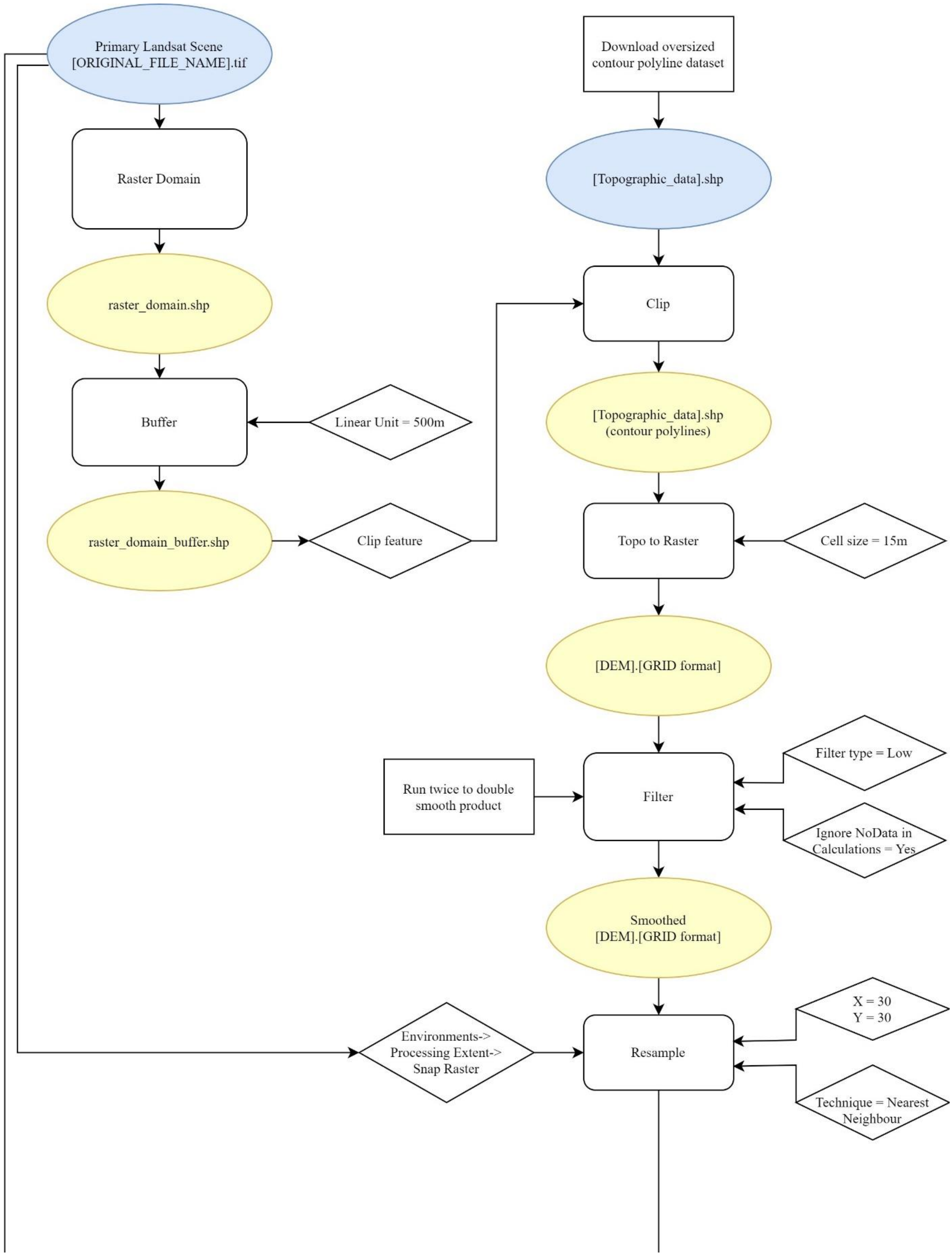
Sample size is a critical variable in any predictive model and machine learning models are known to be 'data hungry' and perform best when provided with a large volume of samples. The classifier model used in this research overall, operated non-parametrically and the structure of the data selected from the primary satellite dataset by the combined ground truth dataset was not assumed to follow any predetermined pattern of distribution (Raschka. 2020). Regardless of this, classes with low representation by area in the combined ground truth dataset were targeted in surveys as the assumption was made that the classifier's ability to detect rare classes would be improved if more sample area was available.

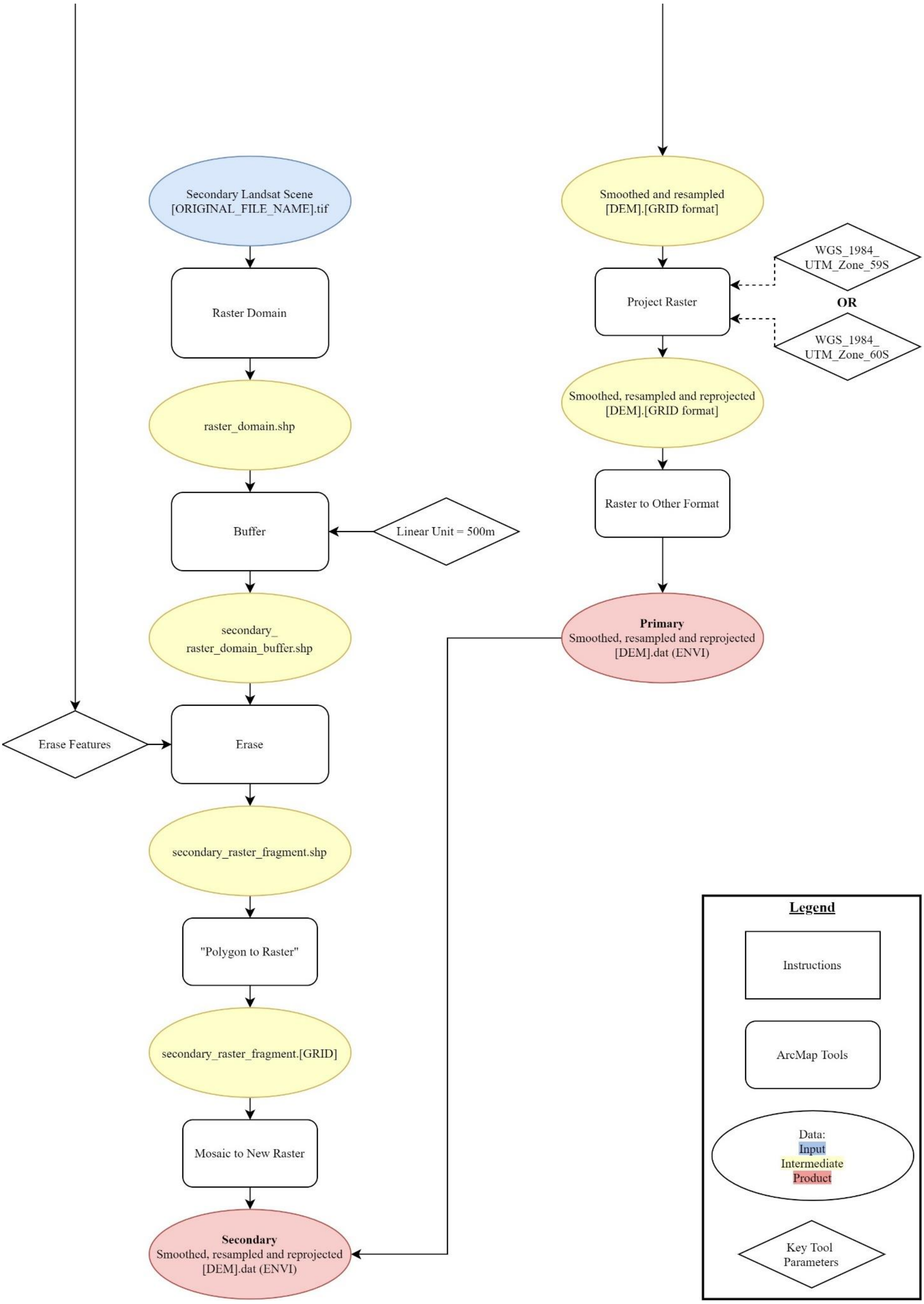
3.3 Workflow B: Digital Elevation Model (DEM) Preparation

The DEM in this research was produced primarily for use in topographic correction of satellite image data (Workflow C1), but also for experimentation in improving classifier performance (Workflow D). For it to be accurate and suitable for use in these tasks, the DEM needed to be free of interpolation artefacts, have a spatial resolution equal to the satellite data, and have a format and boundaries suitable for use with different Landsat scenes in ATCOR 3.

This DEM preparation process was modified from the method described in Shepherd and Dymond (2003). This method produces DEMs at a higher resolution than both the source data (NZTopo50 series contour lines, LINZ, 2019) and the satellite image data through use of a contour line interpolation algorithm. The subsequent low-pass filter smoothing aimed to reduce the impact of any interpolation artefacts (such as anomalous "NoData" pixels) that would have effects on correction and classification processes. Resampling this smoothed data provided an opportunity to ensure that the pixels of the DEM and satellite data aligned (through use of a 'Snap Raster') so that this dataset could be easily inserted into downstream processes.

Figure 11. Workflow B: Digital Elevation Model (DEM) Preparation




Secondary Landsat Scene
[ORIGINAL_FILE_NAME].tif
Raster Domain
raster_domain.shp
Buffer
Linear Unit = 500m
secondary_
raster_domain_buffer.shp
Erase Features
Erase
secondary_raster_fragment.shp
"Polygon to Raster"
secondary_raster_fragment.[GRID]
Mosaic to New Raster
Secondary
Smoothed, resampled and reprojected
[DEM].dat (ENVI)
Smoothed and resampled
[DEM].[GRID format]
Project Raster
WGS_1984_
UTM_Zone_59S
OR
WGS_1984_
UTM_Zone_60S
Smoothed, resampled and reprojected
[DEM].[GRID format]
Raster to Other Format
Primary
Smoothed, resampled and reprojected
[DEM].dat (ENVI)
Legend
Instructions
ArcMap Tools
Data:
Input
Intermediate
Product
Key Tool
Parameters

Sub-workflow: Secondary DEM production

ATCOR 3 has the strict requirement that each pixel in the satellite image to be topographically corrected must have a matching pixel value in the DEM. This meant that any area where the Landsat scene extended past the boundary of the DEM or where the DEM contained a pixel with a "NoData" value must be fixed before the correction algorithm would execute. The boundaries of Landsat 4 and Landsat 8 scenes corrected in Workflow C1 were never perfectly aligned and two key limitations of the ArcMap tool "Topo to Raster" were found which produced DEMs with truncated extents (details of these limitations are discussed in Section 5.3). These problems precipitated the development of the DEM extension method which is an effective although blunt approach to fixing boundary misalignment issues in DEMs for use in topographic correction with ATCOR 3. This simple method may not be suitable in cases where the AOI is close to the extension area as distortions may occur, possible alternative methods are discussed in Section 5.3.

The DEM Product labelled "Primary" in the workflow chart can theoretically be used to correct other Landsat scenes of the same area, but often the Landsat scene extent will not fit entirely within the DEM extent and ATCOR 3 will not be able to execute. To circumvent this limitation the process described by the column on the left side of the previous page (p. 33, Figure 11) was used to resize the DEM to match historic satellite scenes.

"Topo to Raster" Tool

The key tool used in DEM production within ArcMap "Topo to Raster" was computationally intensive and often took a very long time to process. When calculating a DEM on 'Machine 1' (described in Table 3, p. 15) it was not uncommon for processing to take 8-12 hours. When using small cell sizes or large processing extents there is a significant risk that the process will fail after several hours of processing. The amount of RAM is the limiting factor in a computer's ability to interpolate a raster DEM from contour line data but its precise relationship with the size and spatial resolution of the input and output data is unknown. Ensuring that the computer has a large RAM pool and is able to allocate as much of this as possible to ArcMap when executing the "Topo to Raster" tool is the best approach for minimising the risk of this tool's failure and loss of time.

3.4 Workflow C1: Satellite Image Correction

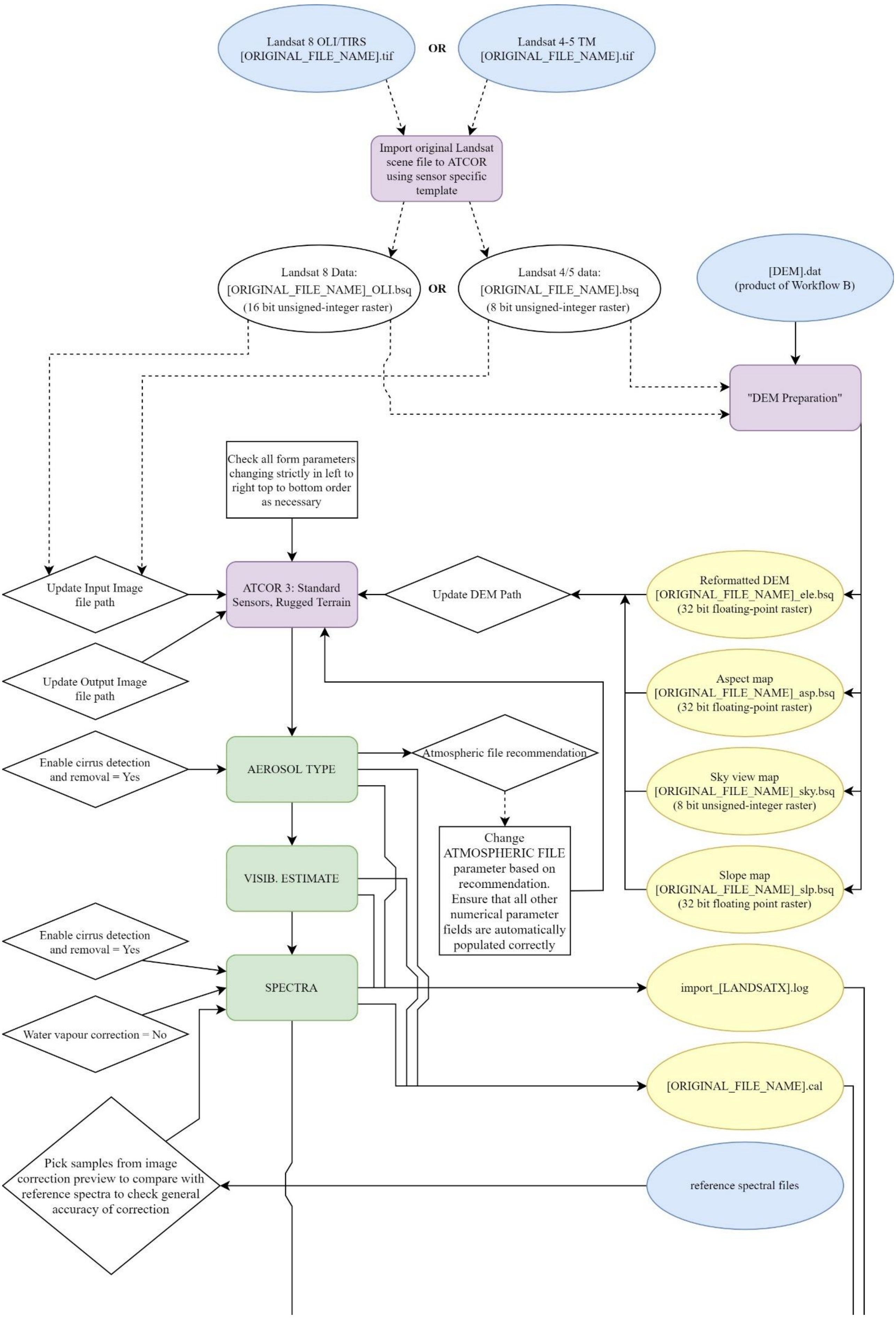
For each of the ATCOR modules described below it is important that the fields are filled out in a left to right, top to bottom order as selection of one parameter may affect the function of fields further on in the form. Adhering to this will reduce the chance of errors and aborts in initialization and execution of this software.

The folder used to store all the input and output data should be new and clean so that file names and paths for the inputs and outputs of each discrete module remain in their default nomenclature. This will facilitate the simple (often automated) transferal of files between each discrete module used in this workflow and will ensure the full processing workflow can be carried out as quickly and efficiently as possible.

Although selecting from the standard sensor modules should automatically populate parameters in the form, it is recommended to check that these values are as expected every time modules are run in case of unexpected changes, bugs or errors. ATCOR 3 workflows for Landsat 8 OLI/TIRS and Landsat 4 TM data differ in their available parameter options, meaning that the generalised workflow described below was slightly different depending on which standard sensor module was applied. Inclusion or exclusion of these parameter options may have a significant impact on the final corrected image values.

The workflow chart on the following page was adapted from the ATCOR-2/3 User Guide, Version 9.3.0 (Richter & Schläpfer. 2019) to provide simplified instructions for using ATCOR 3 to atmospherically and topographically correct Landsat imagery for use in this research project. The vast majority of the functionality of ATCOR 3 is automated, recommendations are given for selection of the MODTRAN atmospheric and aerosol models and much of the ancillary data used to calculate the surface reflectance is provided in the output.

Figure 12. Workflow C1: Image Correction



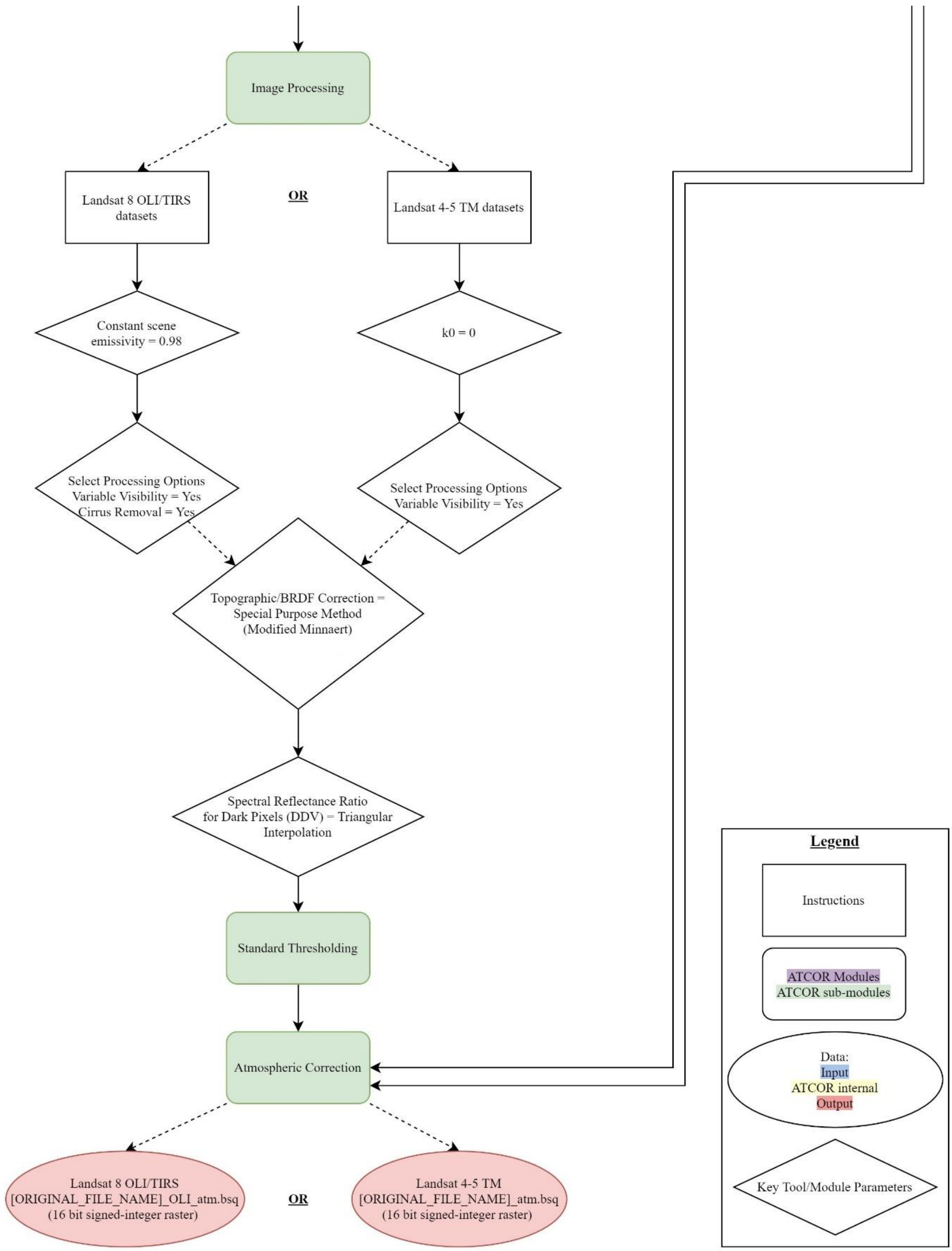


Image Processing
Landsat 8 OLI/TIRS datasets
OR
Landsat 4-5 TM datasets
Constant scene emissivity = 0.98
k0 = 0
Select Processing Options
Variable Visibility = Yes
Cirrus Removal = Yes
Select Processing Options
Variable Visibility = Yes
Topographic/BRDF Correction = Special Purpose Method (Modified Minnaert)
Spectral Reflectance Ratio for Dark Pixels (DDV) = Triangular Interpolation
Standard Thresholding
Atmospheric Correction
Landsat 8 OLI/TIRS
[ORIGINAL_FILE_NAME]_OLI_atm.bsq
(16 bit signed-integer raster)
OR
Landsat 4-5 TM
[ORIGINAL_FILE_NAME]_atm.bsq
(16 bit signed-integer raster)
Legend
Instructions
ATCOR Modules
ATCOR sub-modules
Data:
Input
ATCOR internal
Output
Key Tool/Module Parameters

ATCOR Background

The software package ATCOR 3 was used to carry out atmospheric and topographic image correction. It was selected for its relative simplicity, ease of use, affordability and specialisation in correcting satellite data acquired from rugged terrain. ATCOR software is published by ReSe Applications LLC.

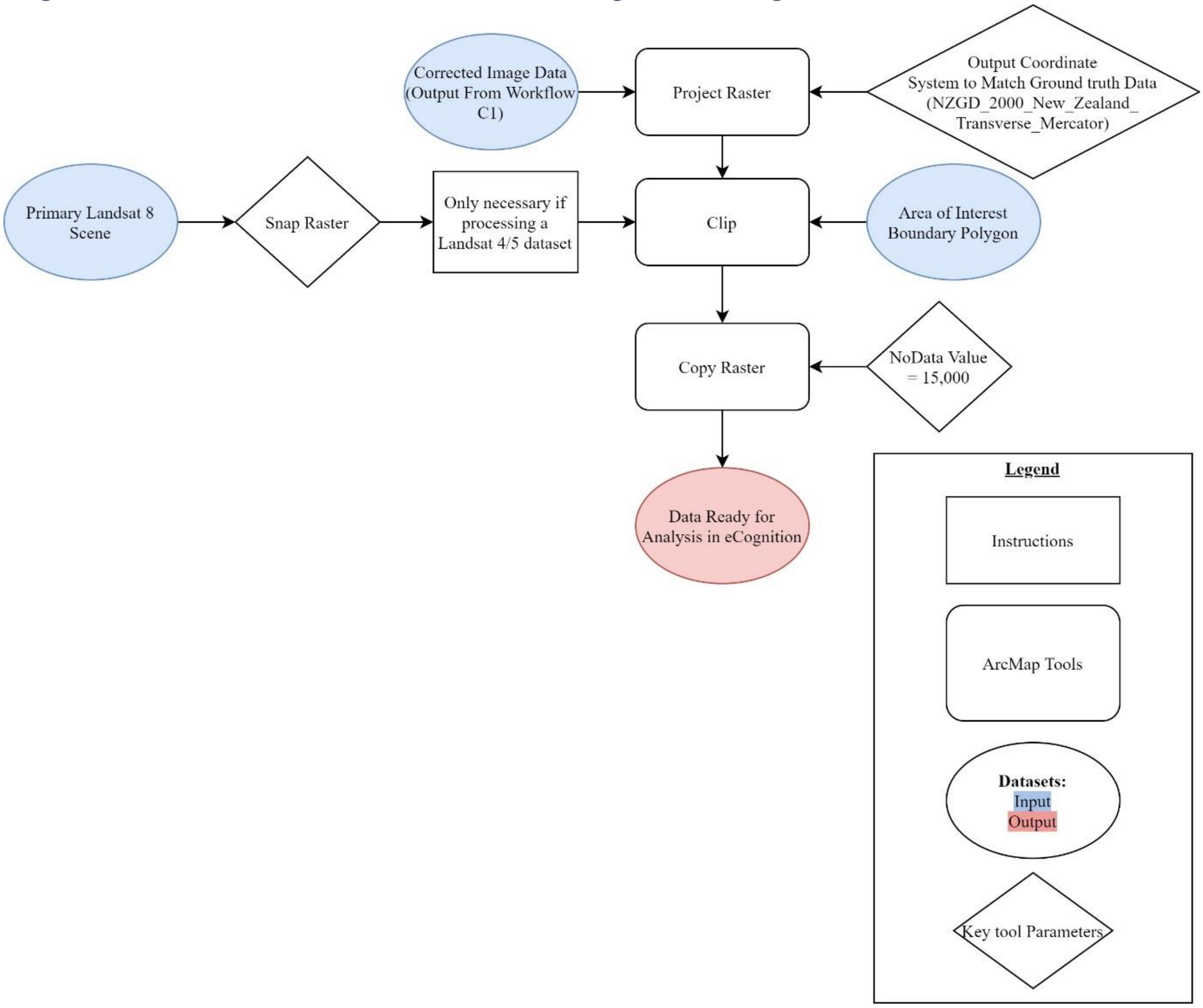
Underlying much of the functionality of ATCOR is the software MODTRAN (MODerate resolution atmospheric TRANsmission), a product developed by Spectral Sciences Inc. and the Air Force Research Laboratory, in its first form, over thirty years ago (Spectral Sciences Inc., n.d.). MODTRAN contains the algorithms for the atmospheric and aerosol models used in the portion of the image correction referred to in this report as "atmospheric correction". These models simulate the physical interactions that affect radiative transfer of electromagnetic energy (light) through the atmosphere as it travels from the Earth's surface to a satellite sensor. The version of ATCOR employed in this research uses the MODTRAN models to supply values to variables necessary for calculating the atmospheric alterations of electromagnetic spectra that are included in Landsat Digital Number (DN) data. These values are largely surface energy outputs and signal degradation or scattering. MODTRAN measures Watts of power at the sensor and estimates how this value was emitted, transmitted and reflected by and through media, often as a function of area of a surface or depth of a gas column. It makes use of various geometric metadata in the Landsat data product to calculate the distance that photons travelled through the gas column (atmosphere) to reach the sensor.

As of 2016 MODTRAN6 is the most recent iteration of this software, but the version of ATCOR used in this research (version 9.3) uses the code of MODTRAN5 (Richter & Schläpfer, 2019) to calculate the lookup tables of key values of radiative transfer.

3.5 Workflow C2: Pre-Classification Image Processing

The flowchart below describes the steps used to prepare corrected Landsat imagery data, as output from ATCOR 3, for analysis with eCognition. All steps in this workflow were carried out within ESRI ArcMap v10.7.1.

Figure 13. Workflow C2: Pre-Classification Image Processing

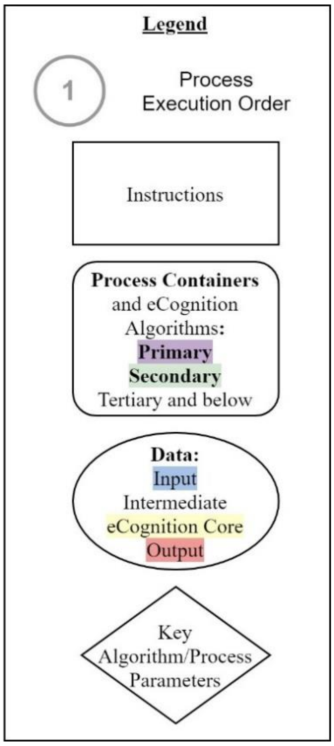


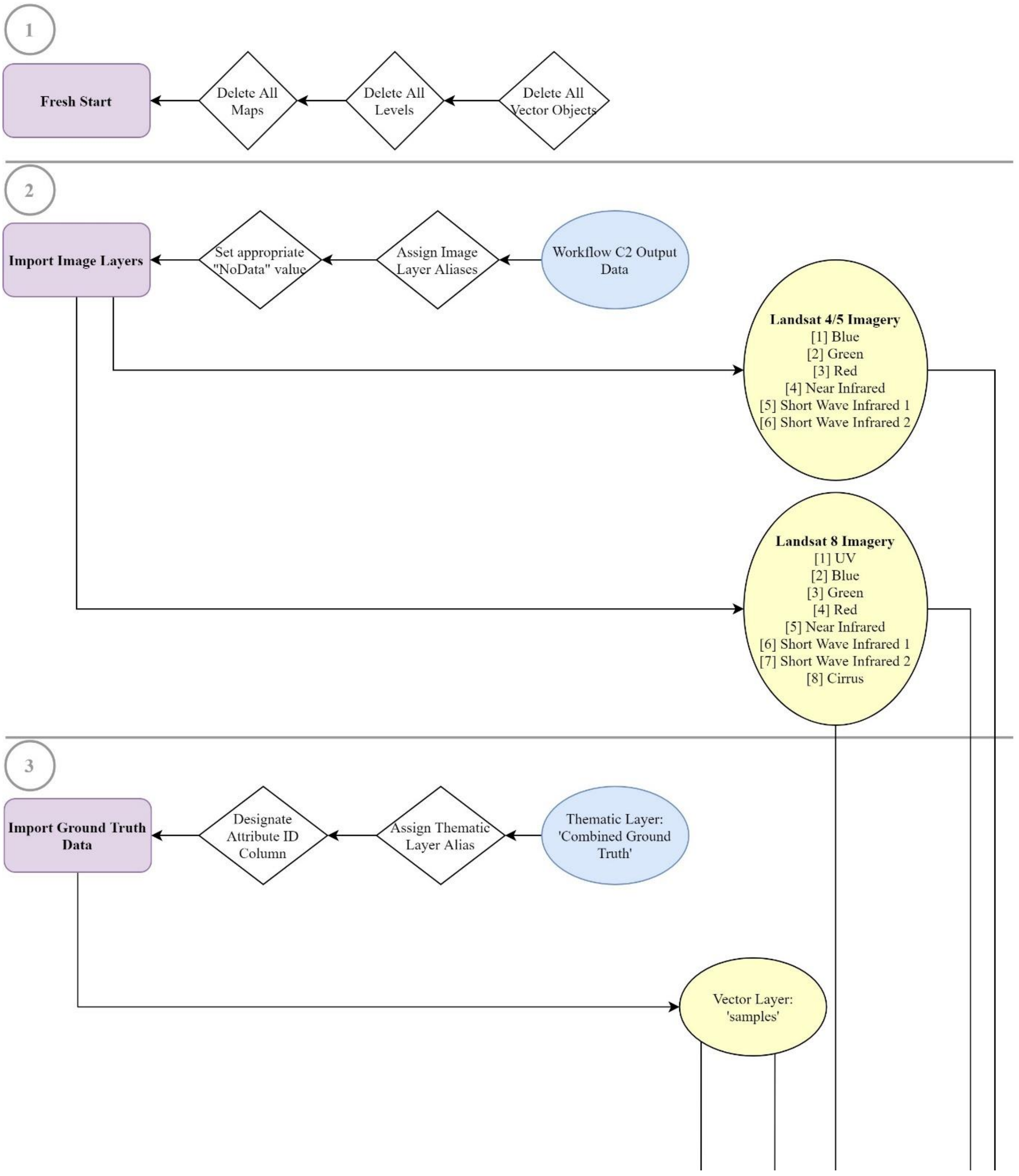
Note. NoData value must be something outside the range of reflectance values output by ATCOR. In this project the value 15000 was used universally to allow for easy, eCognition native exclusion of areas not containing spectral data.

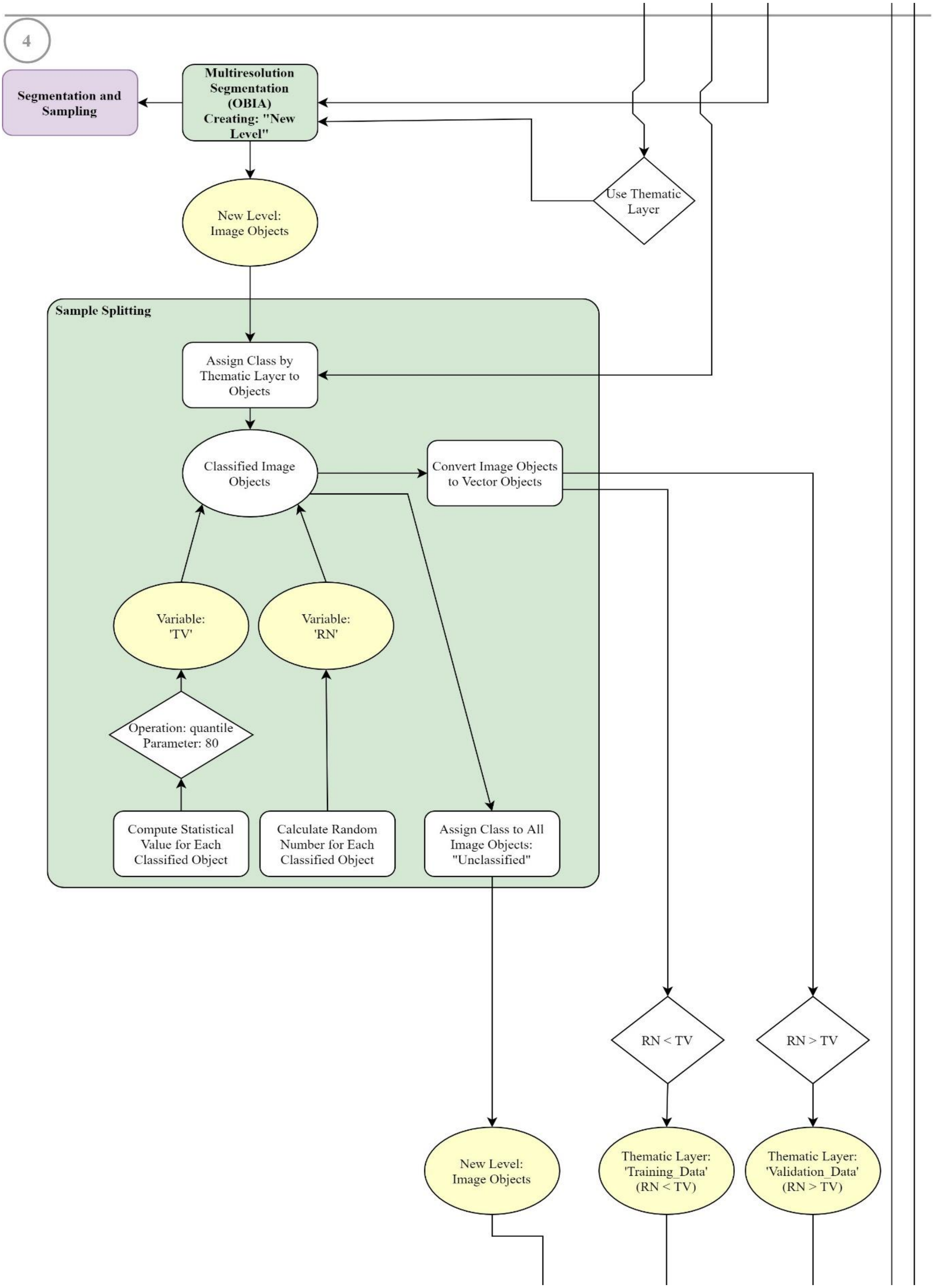
3.6 Workflow D: Automated Land Cover Classification

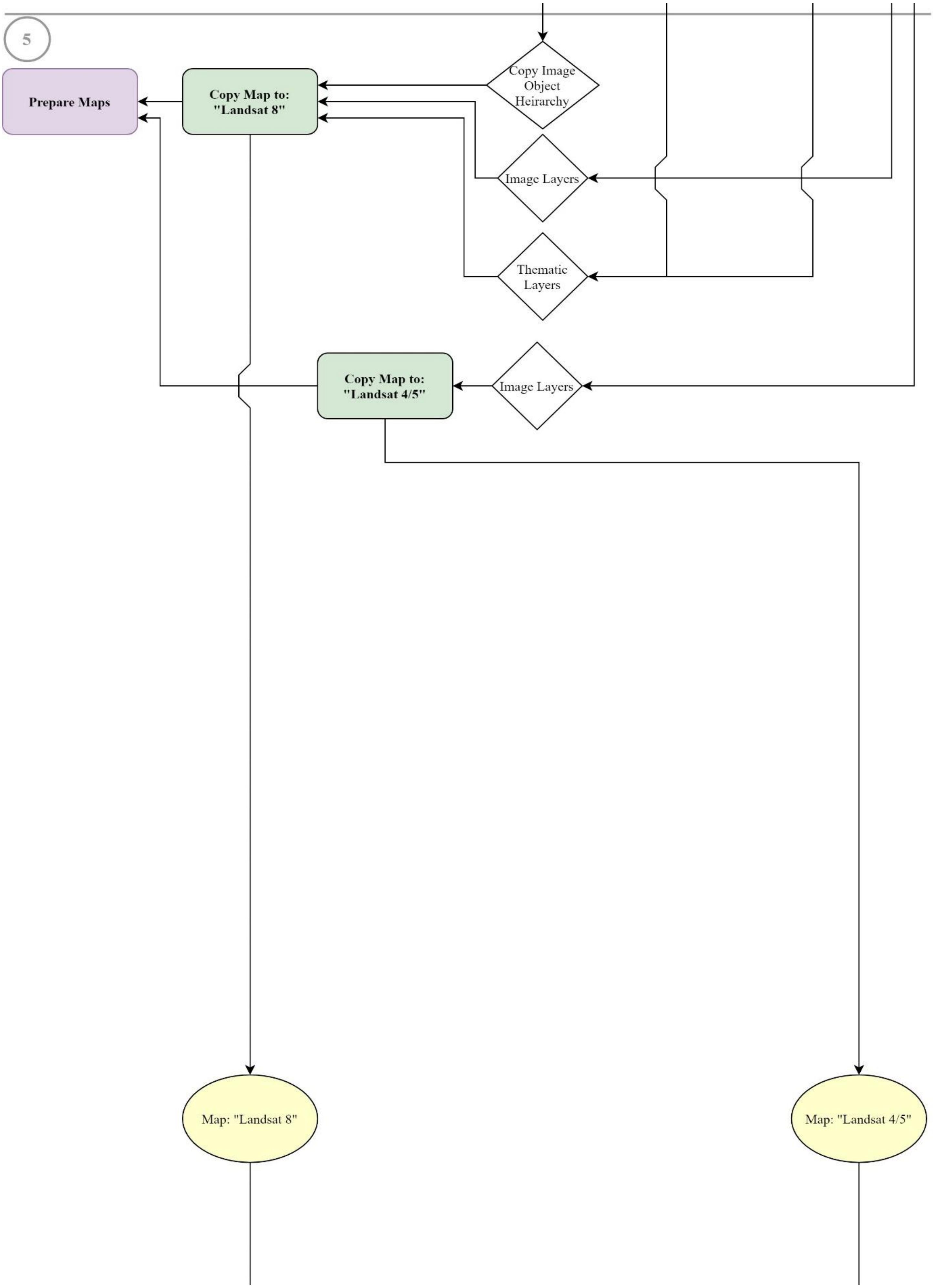
The flowchart presented in this section describes the structure of the data, processes and parameters used to build an automated classifier with Trimble eCognition Developer v9.5.1. All steps in this workflow were carried out within eCognition. This workflow is a simplified transcription of the key parts of the eCognition Process Tree window intended to be used for rapid development of a machine learning classifier for progressing the goals set out in the proposed objectives (Figure 1, p.10)

Figure 14. Workflow D: Automated Classification with Trimble eCognition Developer

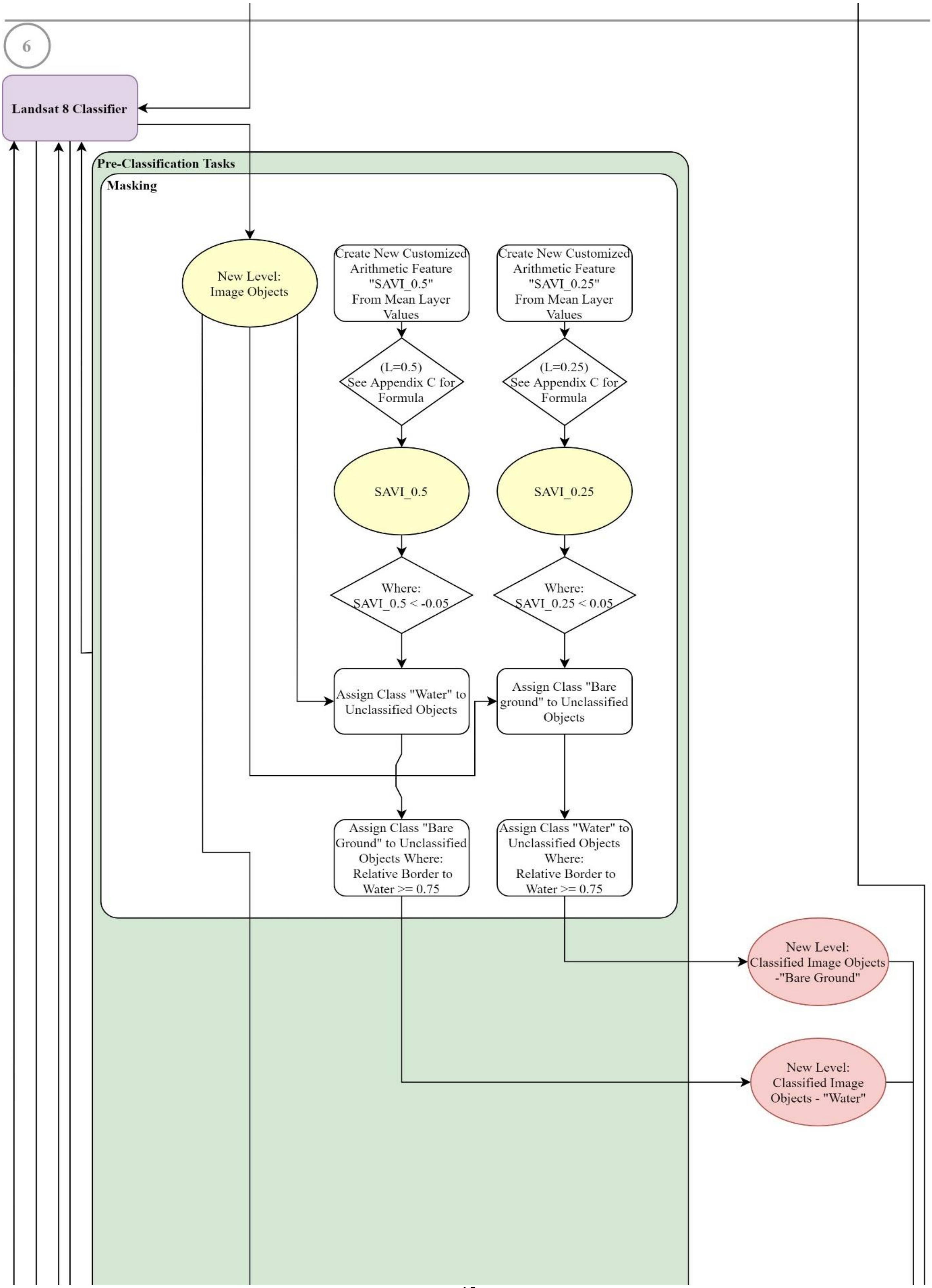


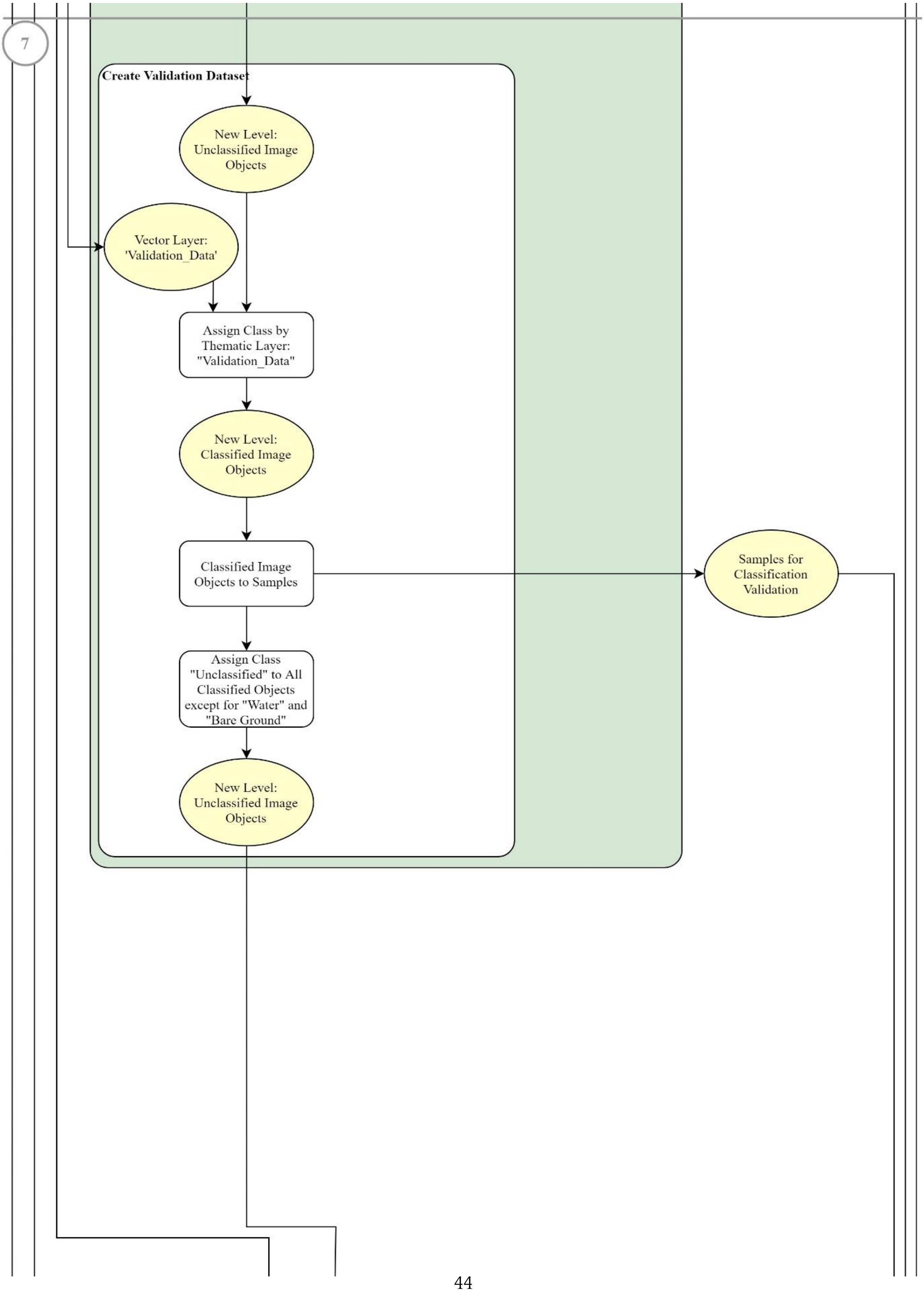

1
Fresh Start
Delete All Maps
Delete All Levels
Delete All Vector Objects
2
Import Image Layers
Set appropriate "NoData" value
Assign Image Layer Aliases
Workflow C2 Output Data
Landsat 4/5 Imagery
[1] Blue
[2] Green
[3] Red
[4] Near Infrared
[5] Short Wave Infrared 1
[6] Short Wave Infrared 2
Landsat 8 Imagery
[1] UV
[2] Blue
[3] Green
[4] Red
[5] Near Infrared
[6] Short Wave Infrared 1
[7] Short Wave Infrared 2
[8] Cirrus
3
Import Ground Truth Data
Designate Attribute ID Column
Assign Thematic Layer Alias
Thematic Layer: 'Combined Ground Truth'
Vector Layer: 'samples'

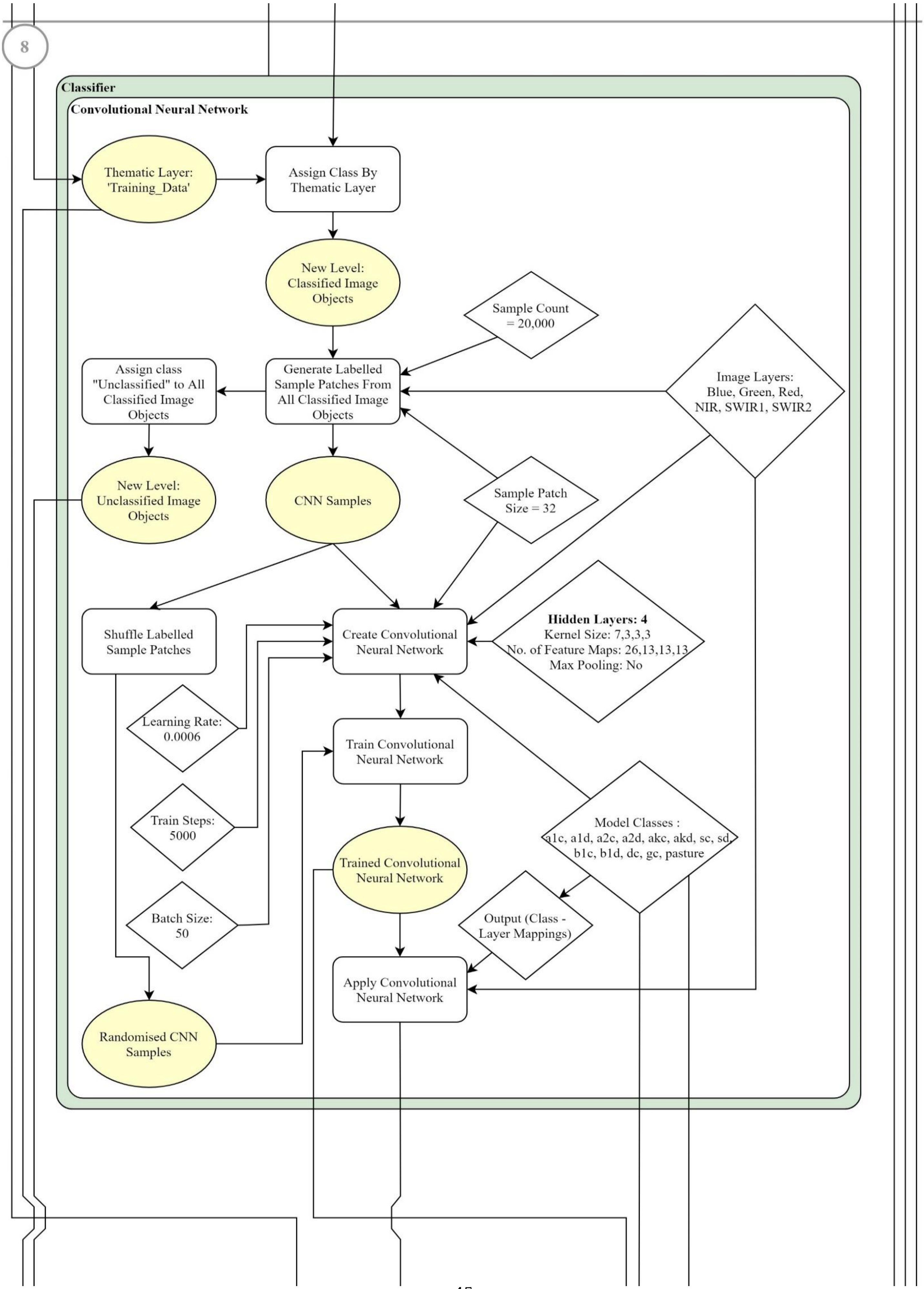

4
Segmentation and Sampling
Multiresolution Segmentation (OBIA) Creating: "New Level"
Use Thematic Layer
New Level: Image Objects
Sample Splitting
Assign Class by Thematic Layer to Objects
Classified Image Objects
Convert Image Objects to Vector Objects
Variable: 'TV'
Variable: 'RN'
Operation: quantile Parameter: 80
Compute Statistical Value for Each Classified Object
Calculate Random Number for Each Classified Object
Assign Class to All Image Objects: "Unclassified"
RN < TV
RN > TV
New Level: Image Objects
Thematic Layer: 'Training_Data' (RN < TV)
Thematic Layer: 'Validation_Data' (RN > TV)

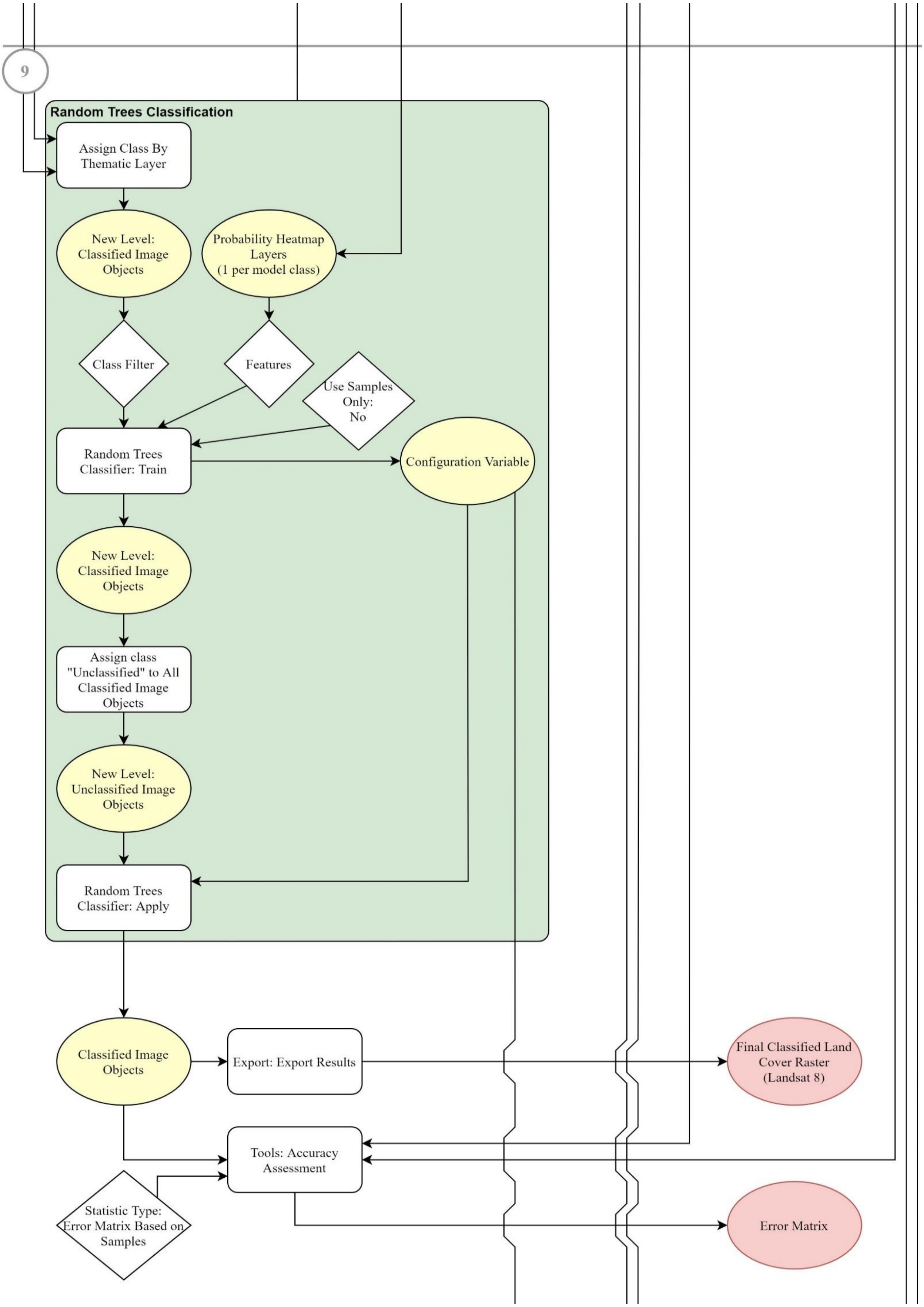


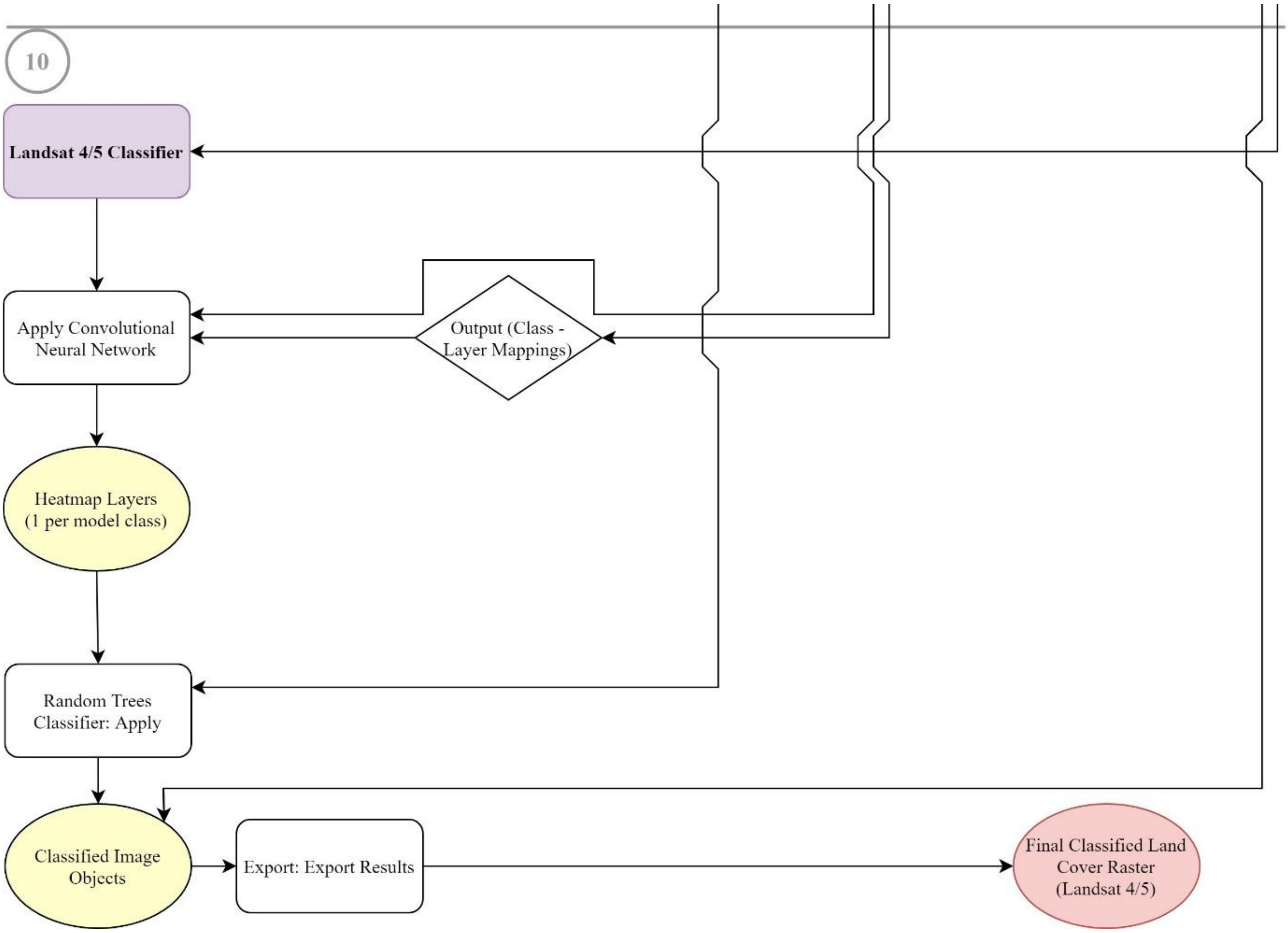
5
Prepare Maps
Copy Map to: "Landsat 8"
Copy Image Object Heirarchy
Image Layers
Thematic Layers
Copy Map to: "Landsat 4/5"
Image Layers
Map: "Landsat 8"
Map: "Landsat 4/5"


6
Landsat 8 Classifier
Pre-Classification Tasks
Masking
New Level: Image Objects
Create New Customized Arithmetic Feature "SAVI_0.5" From Mean Layer Values
Create New Customized Arithmetic Feature "SAVI_0.25" From Mean Layer Values
(L=0.5) See Appendix C for Formula
(L=0.25) See Appendix C for Formula
SAVI_0.5
SAVI_0.25
Where: SAVI_0.5 < -0.05
Where: SAVI_0.25 < 0.05
Assign Class "Water" to Unclassified Objects
Assign Class "Bare ground" to Unclassified Objects
Assign Class "Bare Ground" to Unclassified Objects Where: Relative Border to Water >= 0.75
Assign Class "Water" to Unclassified Objects Where: Relative Border to Water >= 0.75
New Level: Classified Image Objects -"Bare Ground"
New Level: Classified Image Objects - "Water"


7
Create Validation Dataset
New Level: Unclassified Image Objects
Vector Layer: 'Validation_Data'
Assign Class by Thematic Layer: "Validation_Data"
New Level: Classified Image Objects
Classified Image Objects to Samples
Samples for Classification Validation
Assign Class "Unclassified" to All Classified Objects except for "Water" and "Bare Ground"
New Level: Unclassified Image Objects


8
Classifier
Convolutional Neural Network
Thematic Layer: 'Training_Data'
Assign Class By Thematic Layer
New Level: Classified Image Objects
Sample Count = 20,000
Assign class "Unclassified" to All Classified Image Objects
Generate Labelled Sample Patches From All Classified Image Objects
Image Layers: Blue, Green, Red, NIR, SWIR1, SWIR2
New Level: Unclassified Image Objects
CNN Samples
Sample Patch Size = 32
Shuffle Labelled Sample Patches
Create Convolutional Neural Network
Hidden Layers: 4
Kernel Size: 7,3,3,3
No. of Feature Maps: 26,13,13,13
Max Pooling: No
Learning Rate: 0.0006
Train Convolutional Neural Network
Train Steps: 5000
Model Classes : a1c, a1d, a2c, a2d, akc, akd, sc, sd, b1c, b1d, dc, gc, pasture
Trained Convolutional Neural Network
Batch Size: 50
Output (Class - Layer Mappings)
Apply Convolutional Neural Network
Randomised CNN Samples


9
Random Trees Classification
Assign Class By Thematic Layer
New Level: Classified Image Objects
Probability Heatmap Layers (1 per model class)
Class Filter
Features
Use Samples Only: No
Random Trees Classifier: Train
Configuration Variable
New Level: Classified Image Objects
Assign class "Unclassified" to All Classified Image Objects
New Level: Unclassified Image Objects
Random Trees Classifier: Apply
Classified Image Objects
Export: Export Results
Final Classified Land Cover Raster (Landsat 8)
Tools: Accuracy Assessment
Statistic Type: Error Matrix Based on Samples
Error Matrix


10
Landsat 4/5 Classifier
Apply Convolutional Neural Network
Output (Class - Layer Mappings)
Heatmap Layers (1 per model class)
Random Trees Classifier: Apply
Classified Image Objects
Export: Export Results
Final Classified Land Cover Raster (Landsat 4/5)

4. Results

The results of this research are presented as a set of data displays aimed at providing a generalised diagnostic framework for improving methods of sampling, image processing and classification.

4.1 Sampling Protocols

Machine learning classifiers require a large amount of sample data to produce accurate results. Labelled sample datasets used to train classifiers must have strict definitions and be of consistent quality in order to achieve optimal performance from the classifier. The method described in Workflow A (Section 3.2) provides a structure for producing consistent, high quality data appropriate for use with the machine learning classifier described in Workflow D (Section 3.6). The protocols set out in Workflow A are a key outcome of this project where great improvement was made over the research period.

Opportunities to improve classification accuracy through further development of the ground truth data production protocols remain, mainly in the Survey Methods and Classification Framework sections (described on p. 19-20, 29). An optimal compromise between the set of classes most useful for describing environmental features and the set of classes most accurately discriminable from the ground truth data production process was not found in this research. The classification framework used in the analysis of this research is presented in Table 6 page 54.

Figures 15 to 17 show the finalised combined ground truth sample sets used to train the classifier model and assess its accuracy for each AOI in context.

Table 5 (p. 52) compares the sample sizes of the classes from the final land cover classification framework (Table 6, p.54) used to create the combined ground truth datasets for each AOI. This table is intended to be a tool to help improve future ground truth sampling protocols for better classification accuracy. A simple linear regression was carried out to test the assumption that a classes' classification accuracy improves in response to increase in sample size. The results could not confirm this hypothesis, they instead indicate that the observed variation in classification accuracy is the result of a variety of unexplored variables.

Figure 15. Combined Ground Truth Samples Overlaying Primary Satellite Data (Canterbury AOI)

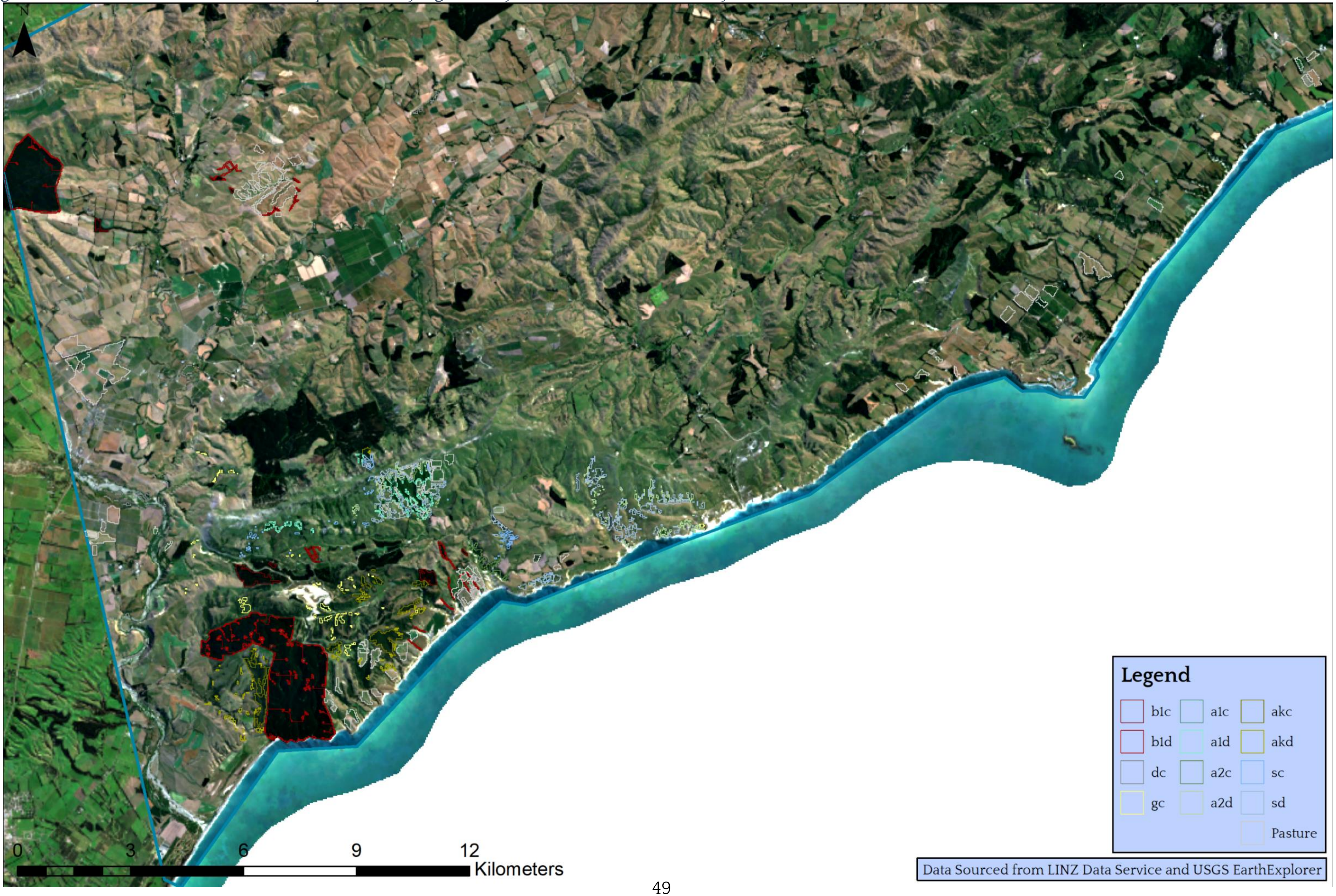


Figure 16. Combined Ground Truth Samples Overlaying Primary Satellite Data (Manawatū-Whanganui AOI)

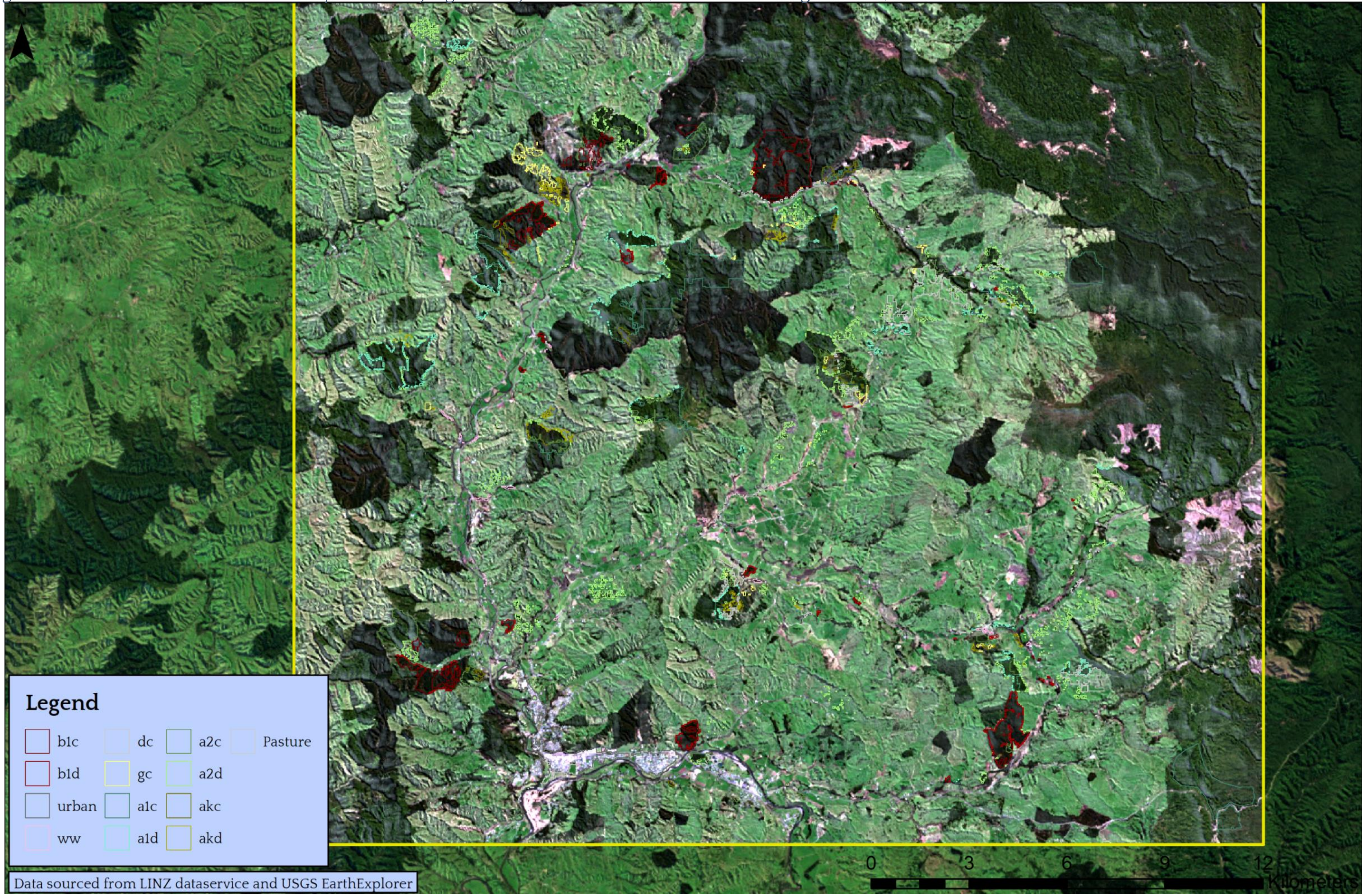


Figure 17. Combined Ground Truth Samples Overlaying Primary Satellite Data (Northland AOI)

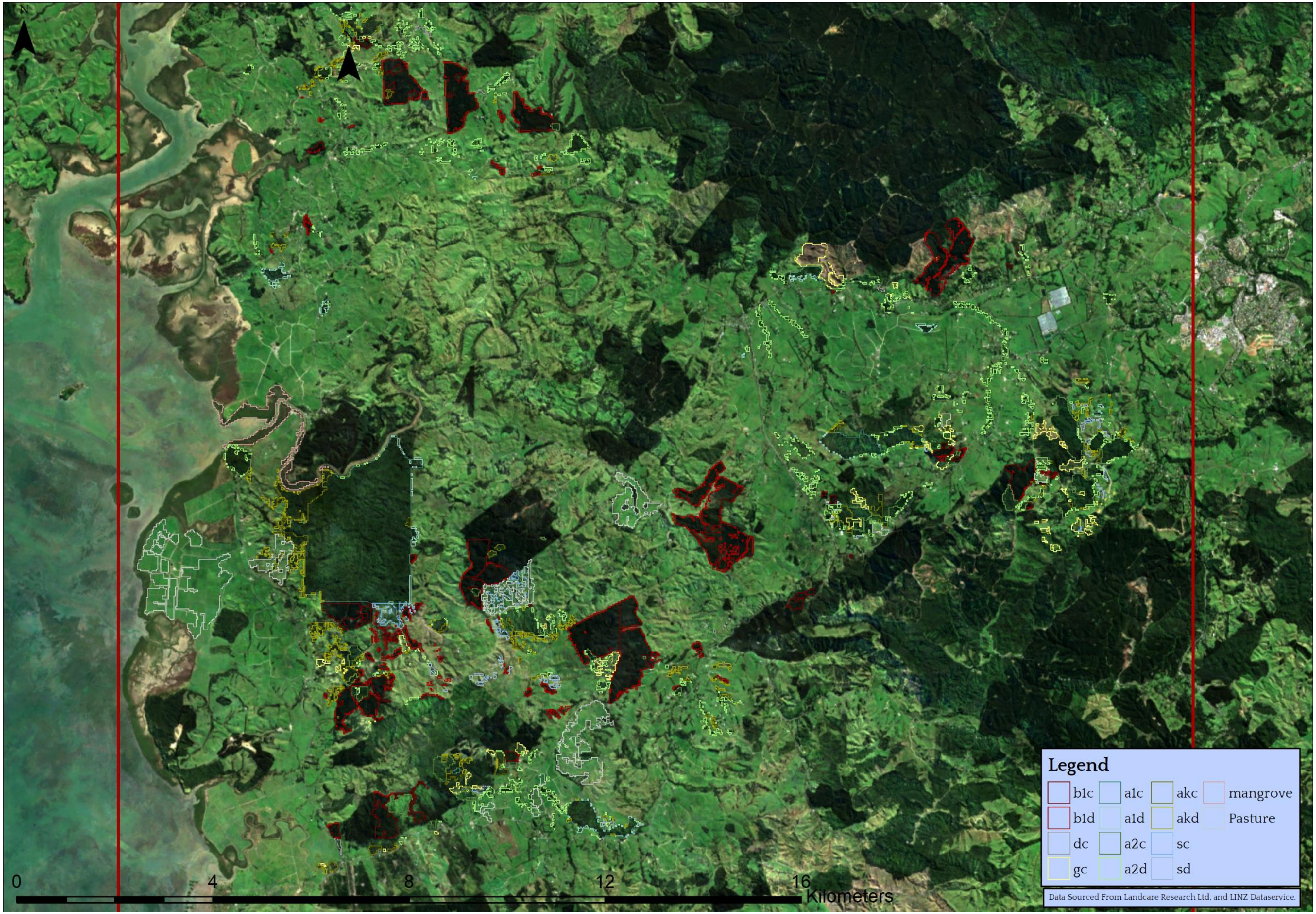


Table 5. Area of Ground Truth Samples and Classified Area of Land Cover Estimated From Primary Satellite Dataset by Class

Class	Canterbury				Manawatū-Wanganui				Northland			
	Ground Truth Sample Area		Classified Area		Ground Truth Sample Area		Classified Area		Ground Truth Sample Area		Classified Area	
	ha	% of AOI	ha	% of AOI	ha	% of AOI	ha	% of AOI	ha	% of AOI	ha	% of AOI
a1c	66.21	0.08	260.55	0.33	1550.56	1.75	16009.47	18.11	709.02	1.10	1412.37	2.18
a1d	22.13	0.03	1329.84	1.71	92.40	0.10	6128.73	6.93	22.96	0.04	2421.90	3.74
a2c	29.01	0.04	558.36	0.72	560.74	0.63	4588.20	5.19	460.67	0.71	5943.15	9.19
a2d	38.74	0.05	2748.15	3.53	260.42	0.29	10233.09	11.58	184.91	0.29	9450.63	14.61
akc	101.16	0.13	2343.69	3.01	163.05	0.18	1658.07	1.88	149.70	0.23	4218.75	6.52
akd	23.43	0.03	4750.29	6.10	39.07	0.04	4274.37	4.84	97.26	0.15	3891.96	6.02
sc	22.04	0.03	3355.83	4.31	-	-	-	-	22.33	0.03	1272.51	1.97
sd	108.06	0.14	2214.09	2.84	-	-	-	-	62.19	0.10	4617.90	7.14
Mangrove	-	-	-	-	-	-	-	-	81.77	0.13	2569.23	3.97
b1c	817.83	1.05	3585.69	4.60	683.78	0.77	5765.67	6.52	767.73	1.19	8964.18	13.68
b1d	63.64	0.08	12107.07	15.54	43.47	0.05	4010.94	4.53	108.26	0.17	4918.50	7.61
Pasture	578.26	0.74	35241.93	45.24	247.09	0.28	14884.38	16.84	389.36	0.60	7647.84	11.83
gc	26.59	0.03	1875.96	2.41	69.35	0.08	4673.34	5.29	114.22	0.18	2339.91	3.62
dc	20.70	0.03	5081.76	6.52	37.79	0.04	8549.01	9.67	5.11	0.01	1294.02	2.00
Urban	-	-	-	-	233.09	0.26	1918.44	2.17	-	-	-	-
ww	-	-	-	-	18.25	0.02	4379.58	4.95	-	-	-	-
Water	*	*	1665.99	2.14	*	*	24.21	0.03	*	*	3540.51	5.47
Bare Ground	*	*	778.14	1.00	*	*	42.30	0.05	*	*	256.68	0.40
Total ground truth sample area	1917.80	2.46	77,897	100	3999.06	4.49	88,392	100	3175.49	4.91	64,672	100

Note. - = Not Applicable in AOI; * = Not Sampled.

See Table 6 for land cover class definitions.

4.2 Classification Framework

The classification framework used in the final iteration of the classification model (Table 6) was selected as it struck a balance between the amount of relevant detail in each class definition and the accuracy of the classified maps produced using it. During the process of testing alternative classification framework structures, the mapping accuracy was found to decrease as number of classes and complexity of class structure increased. A classifier trained with only three classes for example ("Native Vegetation", "Exotic Vegetation" and "Pasture"), had very good accuracy in classifying the primary satellite dataset and produced a somewhat accurate product when applied to secondary satellite data. The amount of detail in this classification framework was however too low for the produced maps to have relevance to the goals of this research.

The final classification framework was built up from the success of the three-class structure, initially through use of only continuous density land cover classes. The final classification framework does not include any sparse density vegetation classes and only includes diffuse density classes where a reasonable visual accuracy could be achieved. Each diffuse class was introduced in turn and the classified maps produced at each iteration were used to visually assess class accuracy. Eventually several diffuse classes were able to be introduced together as confidence in the stability of class definitions grew. Simultaneously the ground truth data was trimmed to ensure that the labelled polygons only included pixels that adhered to the new class structure and definitions. Although the inclusion of diffuse classes was carried out carefully, the simultaneous alteration of polygon boundaries made it impossible to identify which of these two ground truth data parameters was the source of classification error in future maps.

A process diagram in Appendix B provides details of the key changes made to the target classification framework (Appendix A) that lead to the development of the final classification framework (Table 6).

Table 6. Final Land Cover Classification Framework

Code	Class			Present in region	Description
a1c	Native	Remnant	Continuous	C, MW, N	Continuous native woody vegetation with large native Podocarp, Kauri and Angiosperm trees present. Uneven texture to canopy due to high frequency of emergents and complex vertical stratification.
a1d	"	"	Diffuse	C, MW, N	As above but with significant areas of bare ground or pasture.
a2c	"	Regenerating	Continuous	C, MW, N	Higher abundance of native colonising species (Coprosma spp., Cordyline australis, Pittosporum spp., Pseudopanax spp. etc) than in a1 classes, lack of emergent individuals and complexity in the canopy strata.
a2d	"	"	Diffuse	C, MW, N	As above but with significant areas of bare ground or pasture. Large trees tend to have a more sprawling form due to reduced competition for light.
akc	"	Kanuka	Continuous	C, MW, N	Dominated by native Kunzea and/or Leptospermum spp. Remnant/regenerating status can be difficult to assess.
akd	"	"	Diffuse	C, MW, N	As above but with significant areas of bare ground or pasture.
sc	"	Broadleaved Shrubland	Continuous	C, MW, N	Continuous coverage of native broadleaved shrubs (e.g. D. toumatou, Muhlenbeckia spp., Coprosma spp.) and small examples of other tree spp. Most commonly found at high altitude, coastal or disturbed sites.
sd	"	"	Diffuse	C, MW, N	As above but with significant areas of bare ground or pasture.
Mangrove	"	"	"	N	Pixels containing continuous Avicennia marina australasica coverage.
b1c	Exotic	Planted	Continuous	C, MW, N	Planted exotic conifer forest. Common species in this study included: Pinus radiata, Cupressus macrocarpa and Chamaecyparis lawsoniana.
b1d	"	"	Diffuse	C, MW, N	As above but with significant areas of bare ground or pasture. Usually present around edges of plantation blocks or in windbreaks.
Pasture	"	"	Pasture	C, MW, N	Pixels with greater than 85% of the land area covered by Pasture grass or agricultural crop. Also includes native and exotic tussock lands where present.
gc	"	Regenerating	Gorse, continuous	C, MW, N	Pixels with continuous Ulex europaeus or Cystisus spp. coverage.
dc	"	"	Deciduous, continuous	C, MW, N	Primarily Salix and Populus spp. In windbreaks and riparian strips. May be difficult for classifier to discriminate from vineyards.
ww	"	"	Exotic Shrubs, continuous	MW	Regions of continuous exotic shrubs e.g. Rubus spp.
Urban	-	-	Continuous	MW	Pixels covered primarily (> 50%) by buildings, concrete, roads, gardens and other features characteristic of urban environments.
Water	Water	-	Continuous	C, MW, N	Regions of continuous water coverage.
BG	Bare Ground	-	Continuous	C, MW, N	Regions of continuous bare ground or rock.

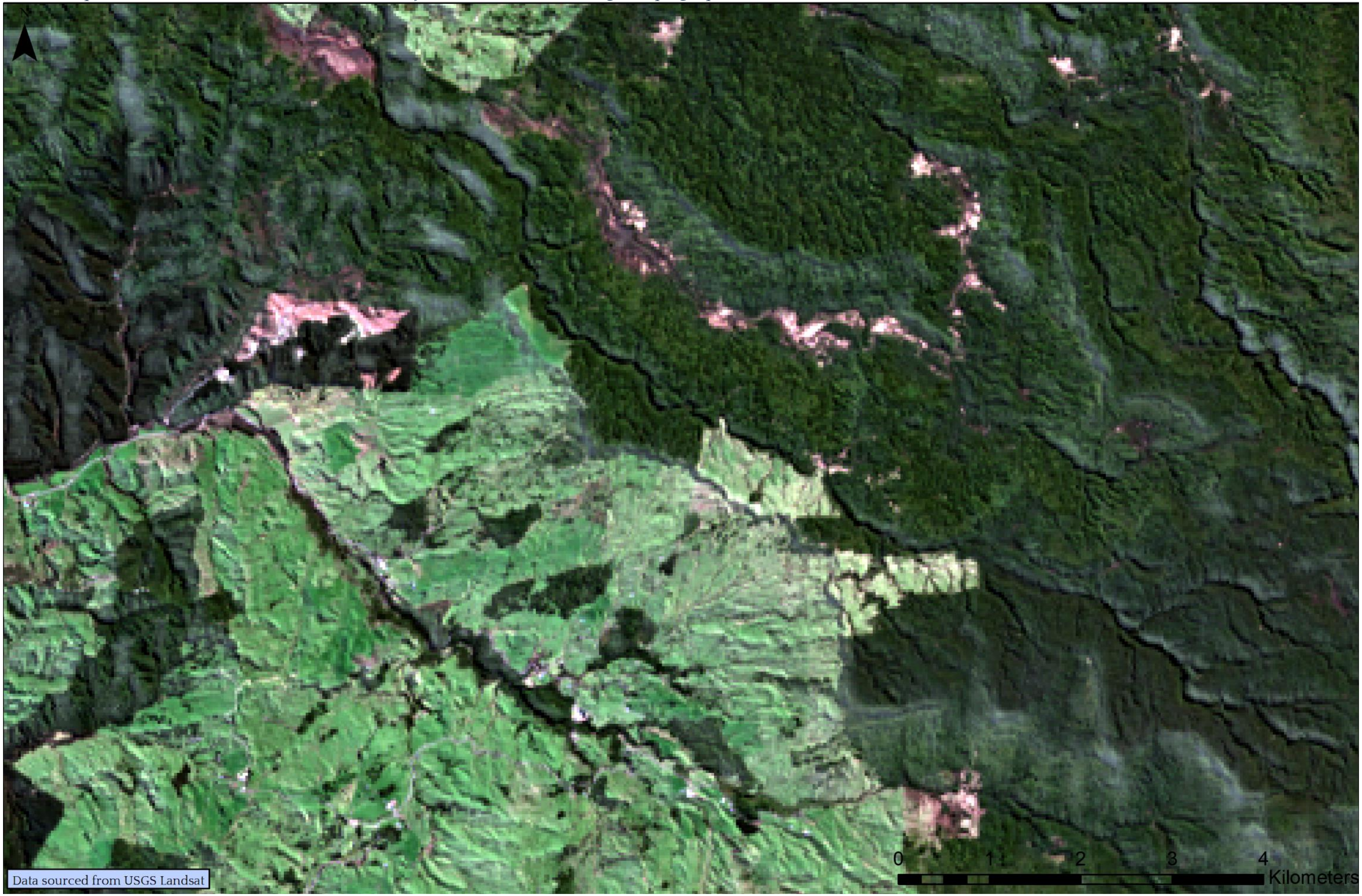
Note. Region Key: C=Canterbury; MW=Manawatū-Wanganui; N=Northland. Continuous classes cover greater than 70% of a pixel's area, diffuse classes cover 15-70% of a pixel's area. Modified from Forbes et al. (2020)

4.3 Data Correction

The chosen corrections model was successful in removing the effects of the atmosphere and topography from the DN Landsat data. However, the topographic correction algorithm produced significant, localised artefacts in the 2016 images of Manawatū-Whanganui and Northland AOIs (Figure 18), these are likely to have contributed to the classification accuracy and image incomparability issues noted in sections 4.6 to 4.9. Overall, the data corrections used did not achieve the goal of normalising the primary and secondary datasets to a standard where the classifier was able to directly compare pixel values with accuracy.

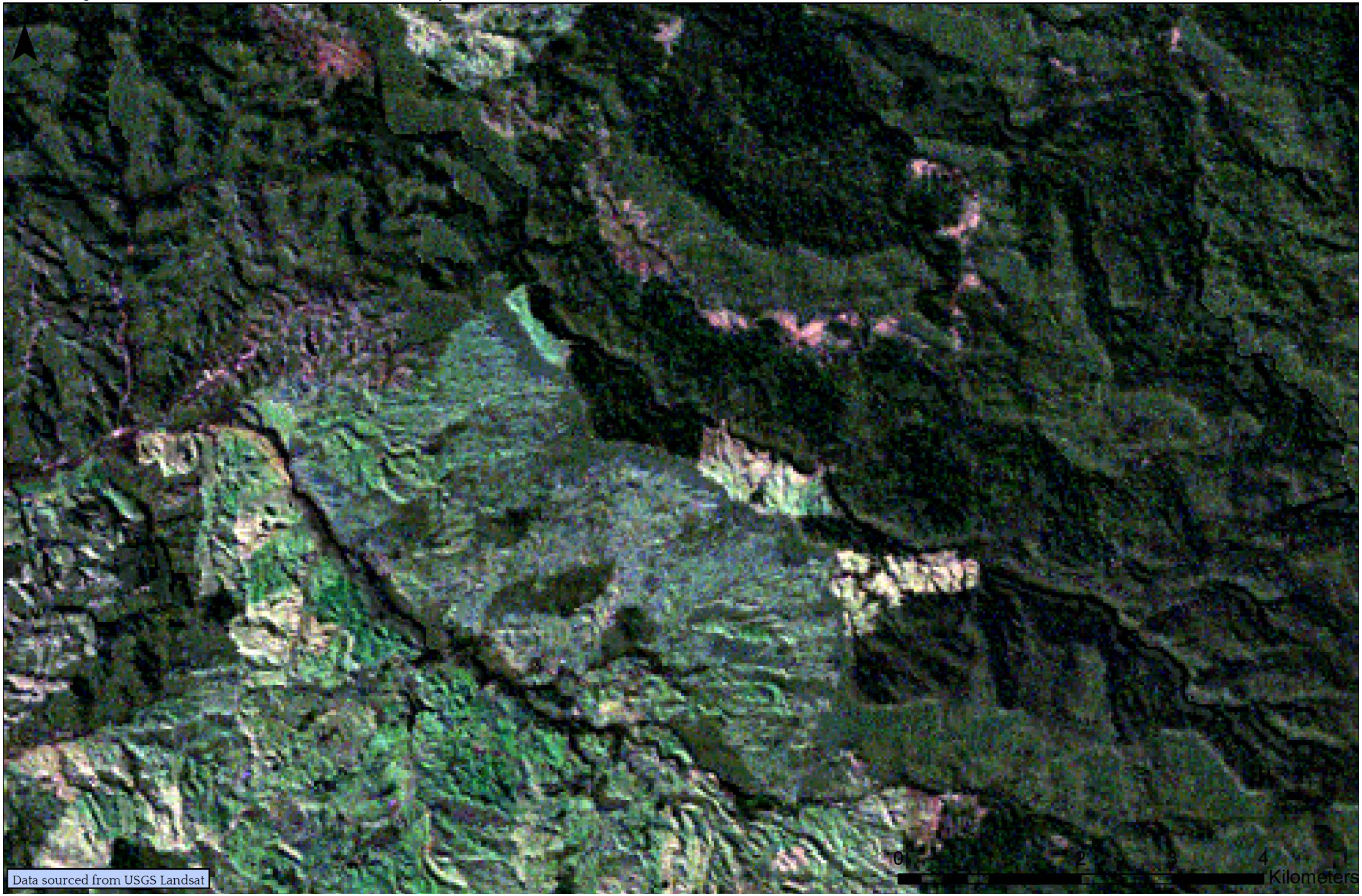
Iterative testing of a number of different satellite data correction methods repeatedly produced similar results suggesting that development of more specific data corrections methods were required. The visually conspicuous data noise seen in areas of uniform colour in all three Landsat 4 images (e.g. Figure 19) is likely to have been a source of error in classification accuracy. A data correction method capable of specifically addressing this artefact pattern is likely to improve classification accuracy in secondary satellite images.

Figure 18. Example of Artefacts Introduced to Primary Satellite Data Through Topographic Correction



Note. Artefacts of topographic correction are visible here as white, cloudlike regions on steep south and west facing slopes of dense forest.

Figure 19. Example of Data Noise Visible in all Secondary Satellite Datasets



Note. Data noise in all Landsat 4 imagery used as secondary satellite data is visible here as random peaks and troughs in pixels intensity over regions of uniform colour.

4.4 Classification Model

The hybrid machine learning classification model produced over the course of this research is specifically set up to analyse 6 channel, medium resolution satellite data through application of ground truth samples in the form of labelled polygon shapefiles. This model's overall accuracy rates were below the generally accepted standard of eighty-five percent classification accuracy (Congalton & Green, 1999) but the wide disparity in the individual accuracy measures of each class indicates these estimates contain error. Important sources of this error likely include:

- The quality of the ground truth data
- The limitations of the classification framework
- The corrected satellite data
- The classifier model itself

Numerous permutations of classifier parameters were tested in the process of developing the optimised model used in this thesis and although these were not tested exhaustively, most improvements in classification accuracy occurred as a result of changes to the input samples (ground truth dataset and classification framework).

4.5 Satellite Datasets and Classified Maps

Over Page.

Note that Water and Bare Ground classes were excluded from analysis as they were defined prior to classifier training using an index threshold of the Soil Adjusted Vegetation Index (see Appendix D and Workflow D).

Figure 20. Corrected Primary Satellite Dataset (Canterbury AOI, Landsat 8, 2014) Superimposed Over Aerial Image Composite

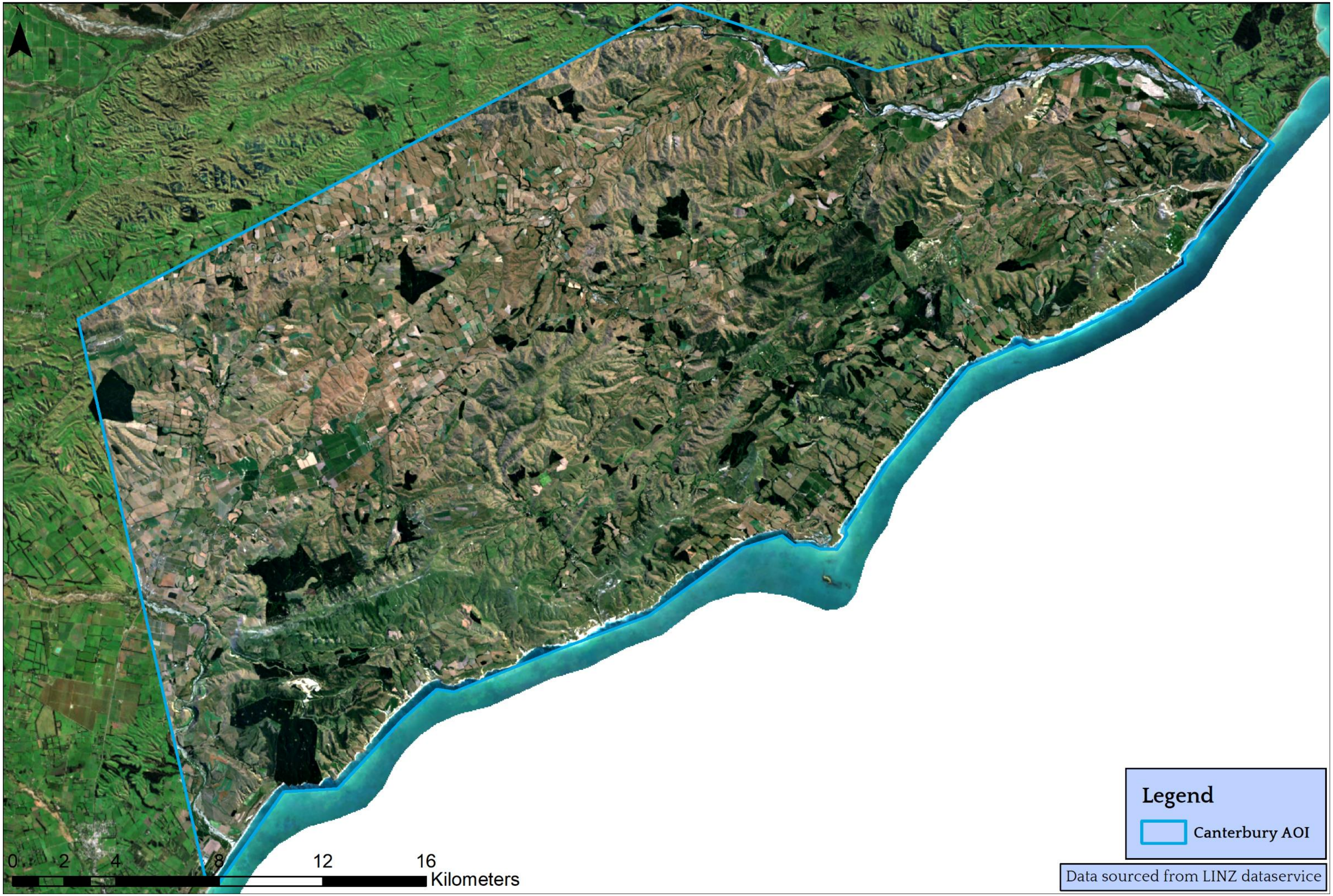


Figure 21. Corrected Secondary Satellite Dataset (Canterbury AOI, Landsat 4, 1990) Superimposed Over Aerial Image Composite

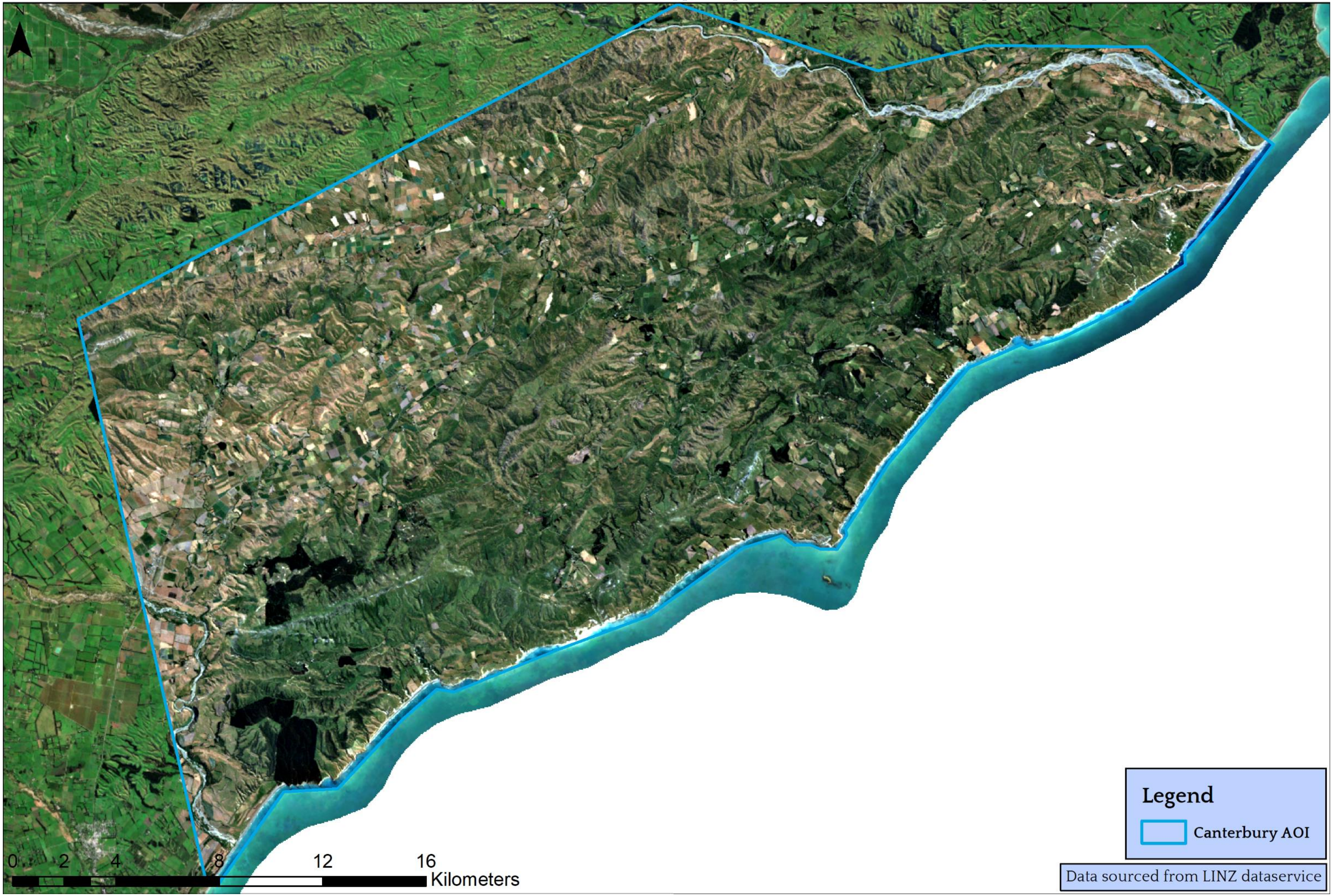


Figure 22. Automatically Classified Land Cover Map of Canterbury AOI (2014)

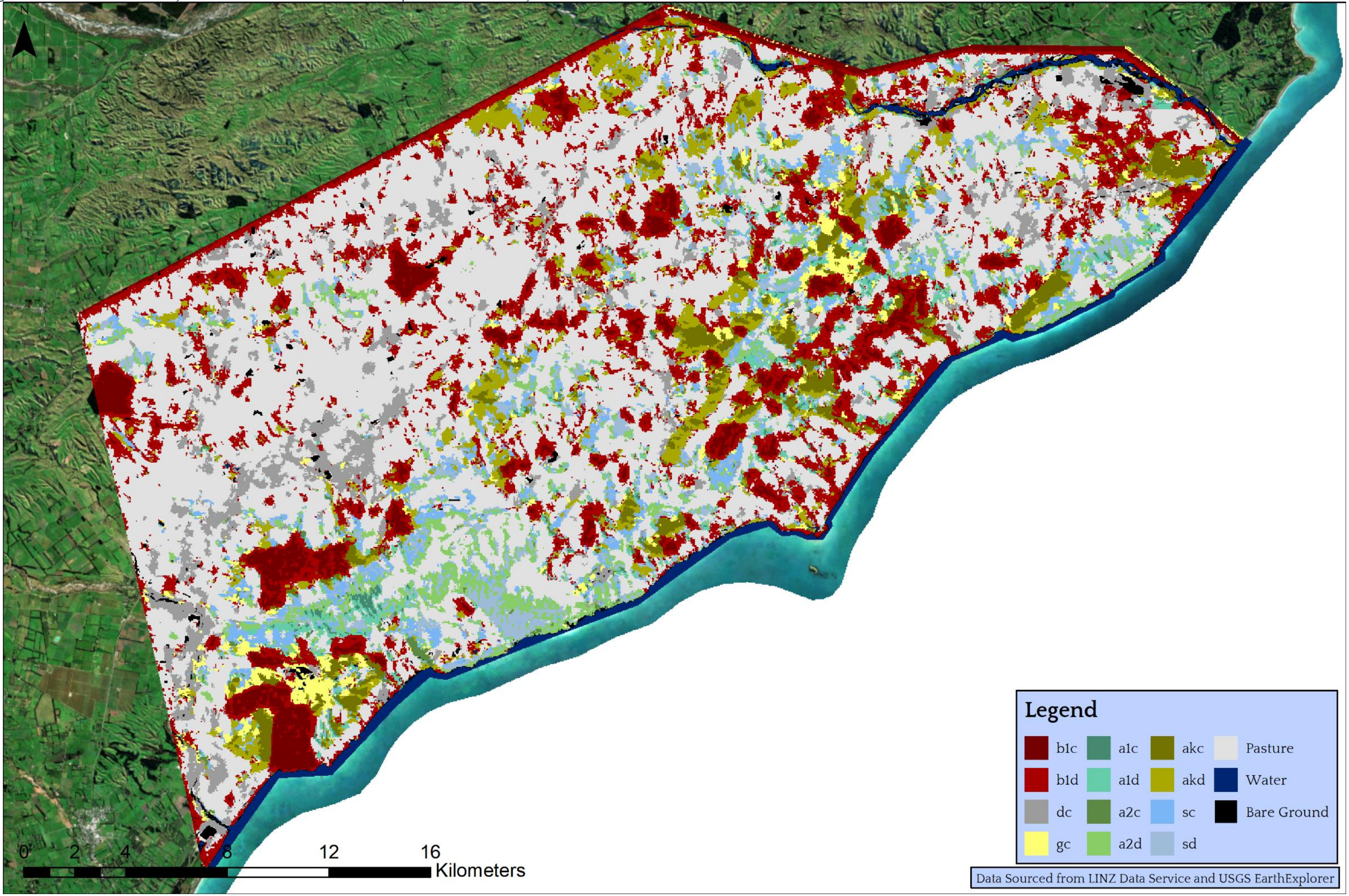


Figure 23. Automatically Classified Land Cover Map of Canterbury AOI (1990)

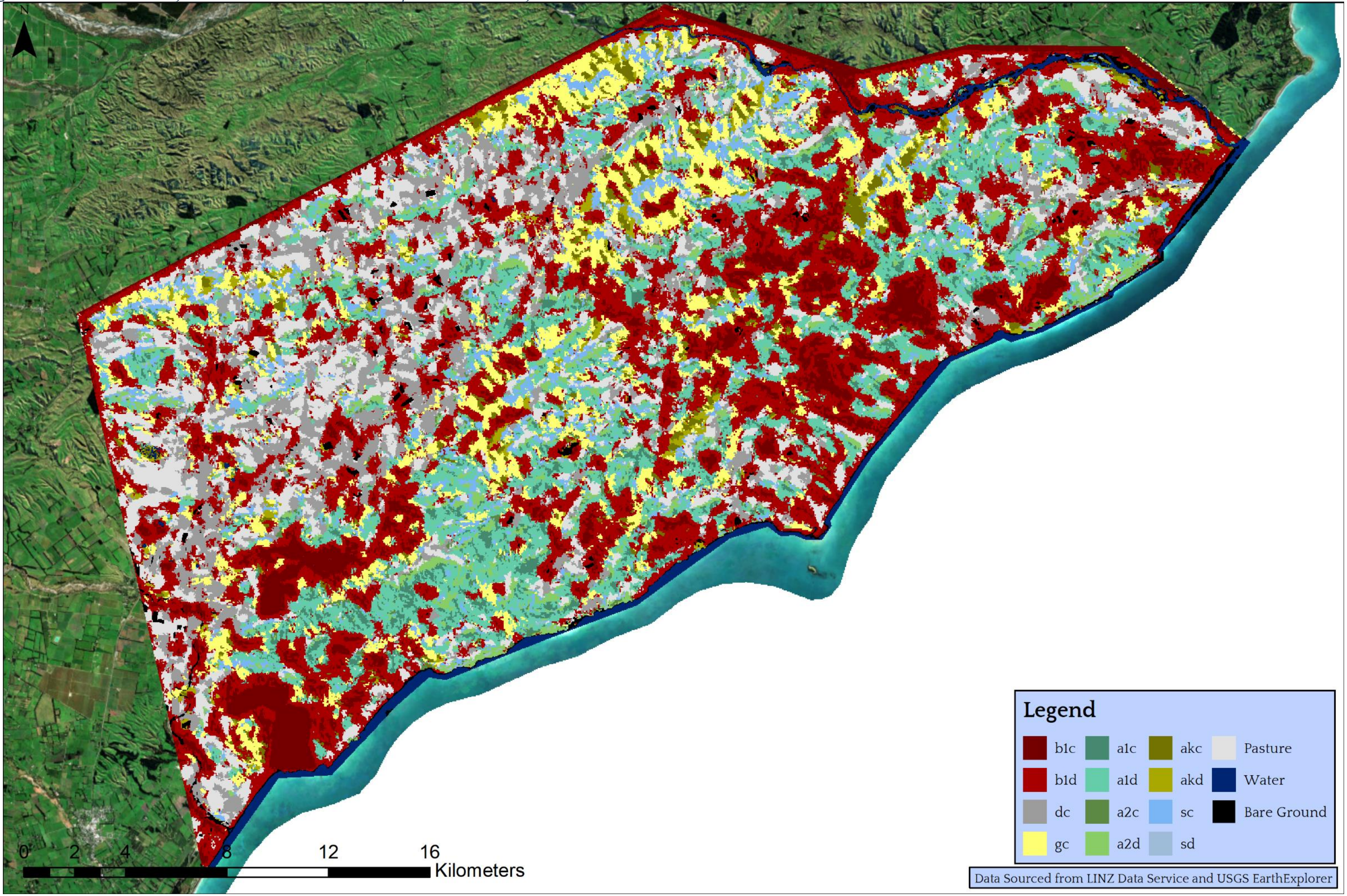


Table 7. Error Matrix of Canterbury AOI Classification (2014) for Accuracy Assessment

User Class \ Sample	b1c	b1d	Pasture	dc	akd	gc	a2d	sd	a2c	sc	akc	a1d	a1c	Sum
b1c	263	27	0	0	0	0	0	0	0	0	0	0	0	290
b1d	73	111	23	1	1	2	0	0	0	0	0	0	0	211
Pasture	1	8	430	4	7	0	3	2	0	2	0	0	0	457
dc	0	4	4	51	0	0	0	0	1	0	0	0	0	60
akd	0	0	4	0	33	0	0	1	0	0	7	0	0	45
gc	1	3	0	0	2	27	0	0	1	1	0	0	0	35
a2d	0	0	9	0	0	2	53	47	8	4	0	1	4	128
sd	0	0	5	0	0	0	16	109	4	4	0	2	0	140
a2c	0	0	0	0	0	1	6	3	31	0	0	0	0	41
sc	0	0	1	0	0	0	7	10	0	31	0	0	0	49
akc	0	0	0	0	8	1	0	0	0	0	54	0	0	63
a1d	0	0	26	0	0	0	4	16	0	1	0	23	27	97
a1c	0	0	1	0	0	0	1	1	2	0	0	8	41	54
unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sum	338	153	503	56	51	33	90	189	47	43	61	34	72	
Producer	0.78	0.73	0.85	0.91	0.65	0.82	0.59	0.58	0.66	0.72	0.89	0.68	0.57	
User	0.91	0.53	0.94	0.85	0.73	0.77	0.41	0.78	0.76	0.63	0.86	0.24	0.76	
Hellden	0.84	0.61	0.90	0.88	0.69	0.79	0.49	0.66	0.70	0.67	0.87	0.35	0.65	
Short	0.72	0.44	0.81	0.78	0.52	0.66	0.32	0.50	0.54	0.51	0.77	0.21	0.48	
KIA Per Class	0.73	0.69	0.80	0.91	0.64	0.81	0.55	0.54	0.65	0.71	0.88	0.66	0.56	
Overall Accuracy	0.75													
KIA	0.71													
Sample Area (ha)	817.83	63.64	578.26	20.70	23.43	26.59	38.74	108.06	29.01	22.04	101.16	22.13	66.21	
Sample Area as % of AOI	1.05	0.08	0.74	0.03	0.03	0.03	0.05	0.14	0.04	0.03	0.13	0.03	0.08	

Figure 24. Producer and User Accuracy Spread with KIA of Each Class (Canterbury AOI)

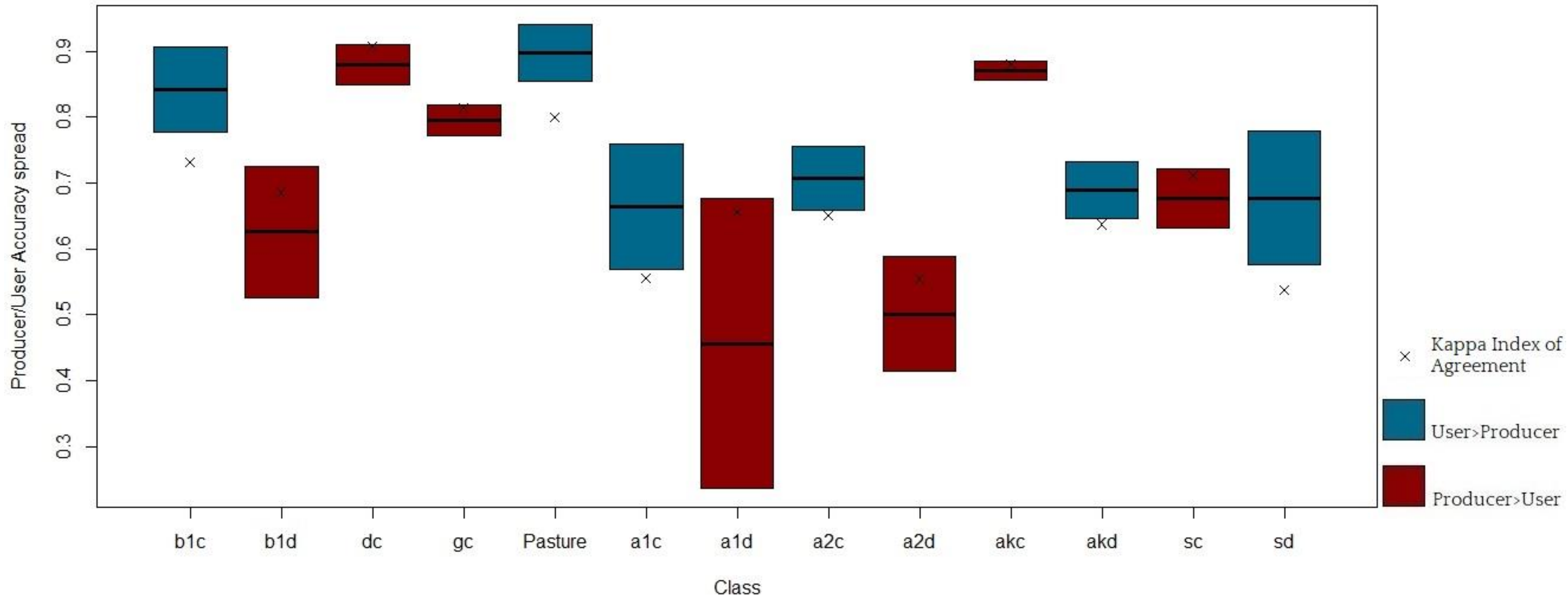
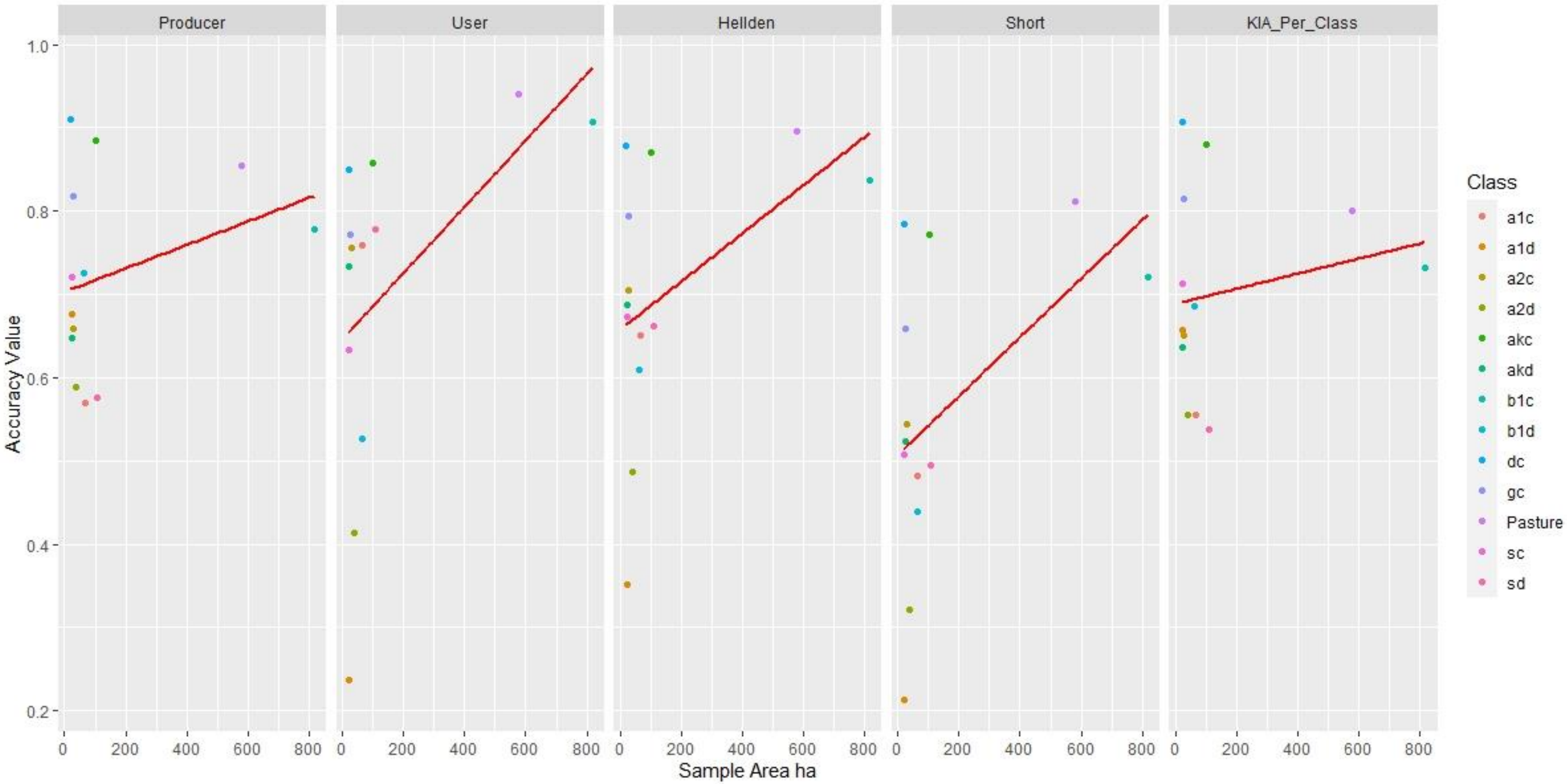


Figure 25. Simple Linear Regression Comparing Class Accuracy Measures With Their Absolute Sample Areas (Canterbury AOI)



4.6 Analysis of Automated Classification (Canterbury AOI)

Imagery

Both images high quality.

Classification Framework

The framework used in classification of this AOI included the eleven core classes as well as continuous and diffuse native broadleaved shrubland.

Land Cover Change

The classifier, trained on primary satellite imagery, was unable to recognise the land cover classes from the final classification framework when applied to the secondary satellite imagery. Estimates of land cover change are of little use as it is obvious on visual inspection of the 1990 land cover map (Figure 23, p. 62) that the accuracy of the predicted data is very low. No reliable changes in land cover are observable in these data.

Accuracy of 2014 classification

Accuracy varies widely between classes but those defined as continuous in the classification framework were generally more accurately classified than those defined as diffuse.

The four classes that were consistently the most accurately classified across all metrics are: continuous planted exotic, continuous deciduous, continuous pasture, continuous kanuka/manuka and continuous gorse.

The four classes that were consistently the least accurately classified classes across all accuracy metrics are: diffuse native remnant, diffuse native regenerating, diffuse native broadleaved shrubland and continuous native broadleaved shrubland.

This AOI was the most accurately classified of the three in this study. As overall accuracy is 0.75 and KIA is 0.71, accuracy should be considered moderate (Congalton and Green, 1999) and it should be assumed that there are still significant errors in the methods used to produce these maps.

Effect of Ground Truth Sample Size on Classification Accuracy

A linear model was fit to each of the five classification accuracy measures to test the effect of sample size on classification accuracy (figure 25, p. 65). Each regression showed a general positive trend with high indicators of significance. R^2 values were poor however so all accuracy assessment values were pooled to test the overall relationship (n=65). This model's fit was also poor (R^2=0.1374).

These simple linear models do not explain enough of the variance in the effect of ground truth sample area on classification accuracy to be able to have confidence in the relationship demonstrated in the regression lines in Figure 25. The high residual standard error and low R^2 values indicate that there are likely other undescribed variables that strongly affect the accuracy of this classifier.

Figure 26. Corrected Primary Satellite Dataset (Manawatū-Whanganui AOI, Landsat 8, 2016) Superimposed Over Aerial Image Composite

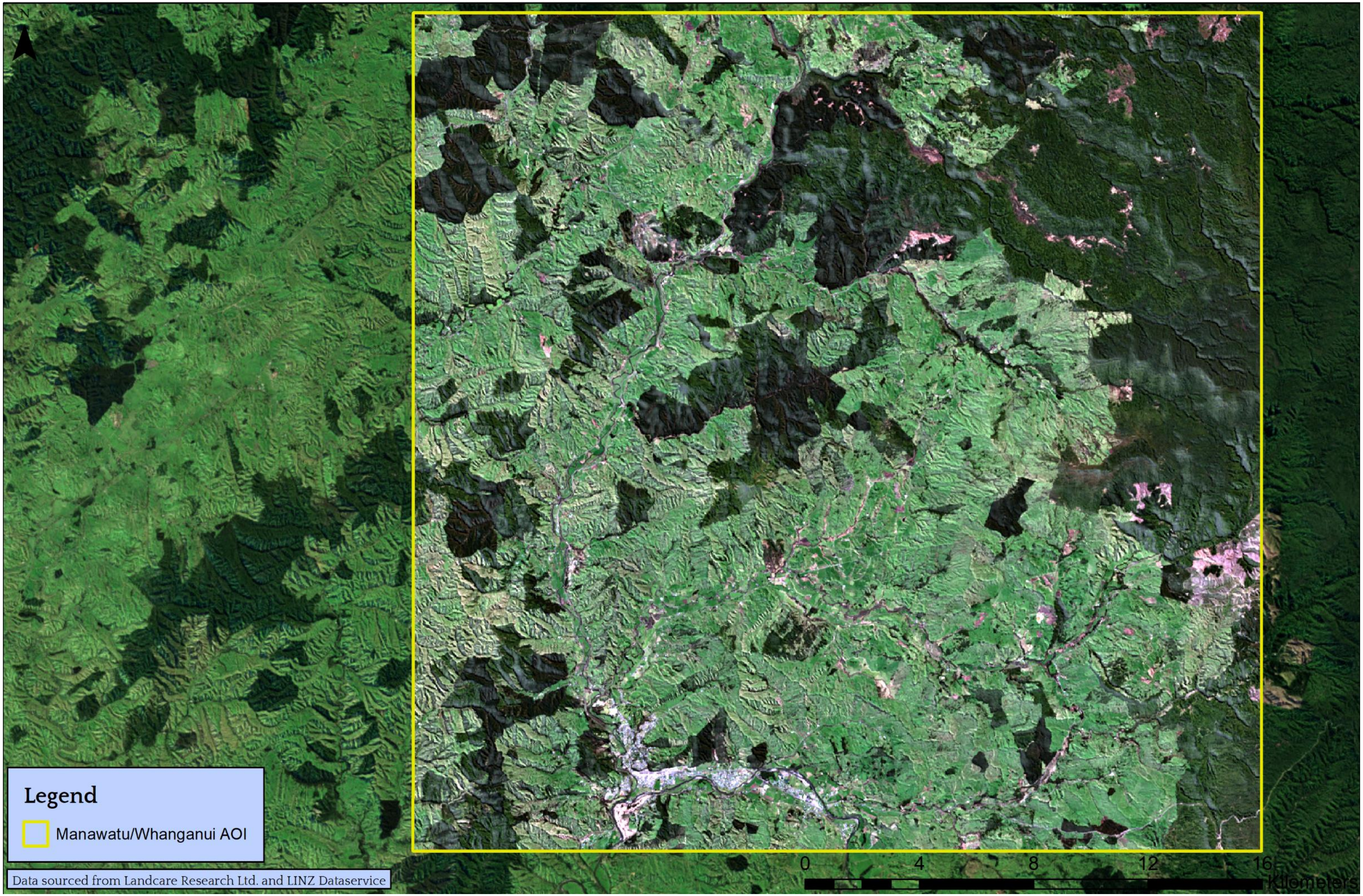


Figure 27. Corrected Secondary Satellite Dataset (Manawatū-Whanganui AOI, Landsat 4, 1989) Superimposed Over Aerial Image Composite

Figure 28. Automatically Classified Land Cover Map of Manawatū-Whanganui AOI (2016)

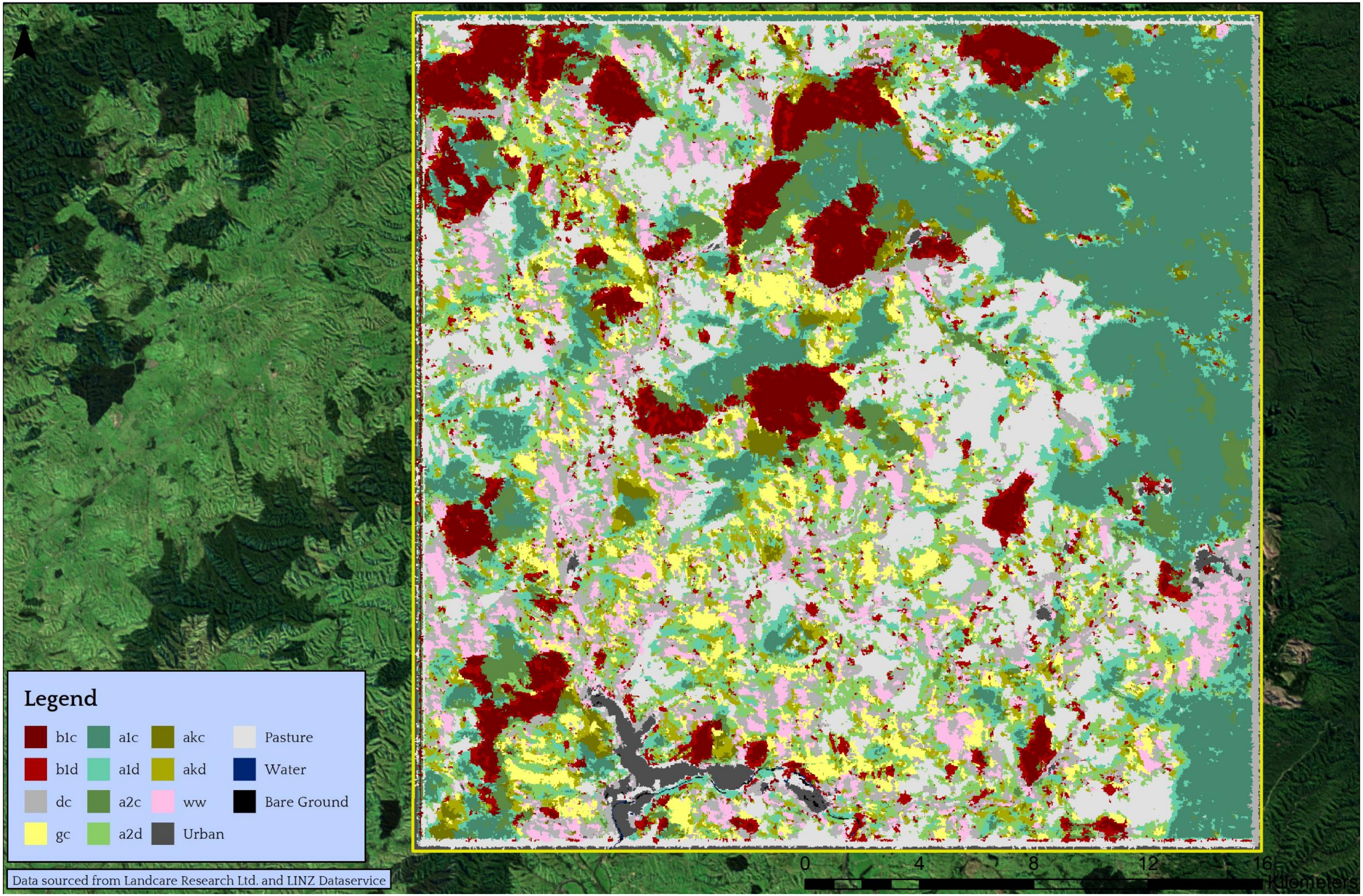


Figure 29. Automatically Classified Land Cover Map of Manawatū-Whanganui AOI (1989)

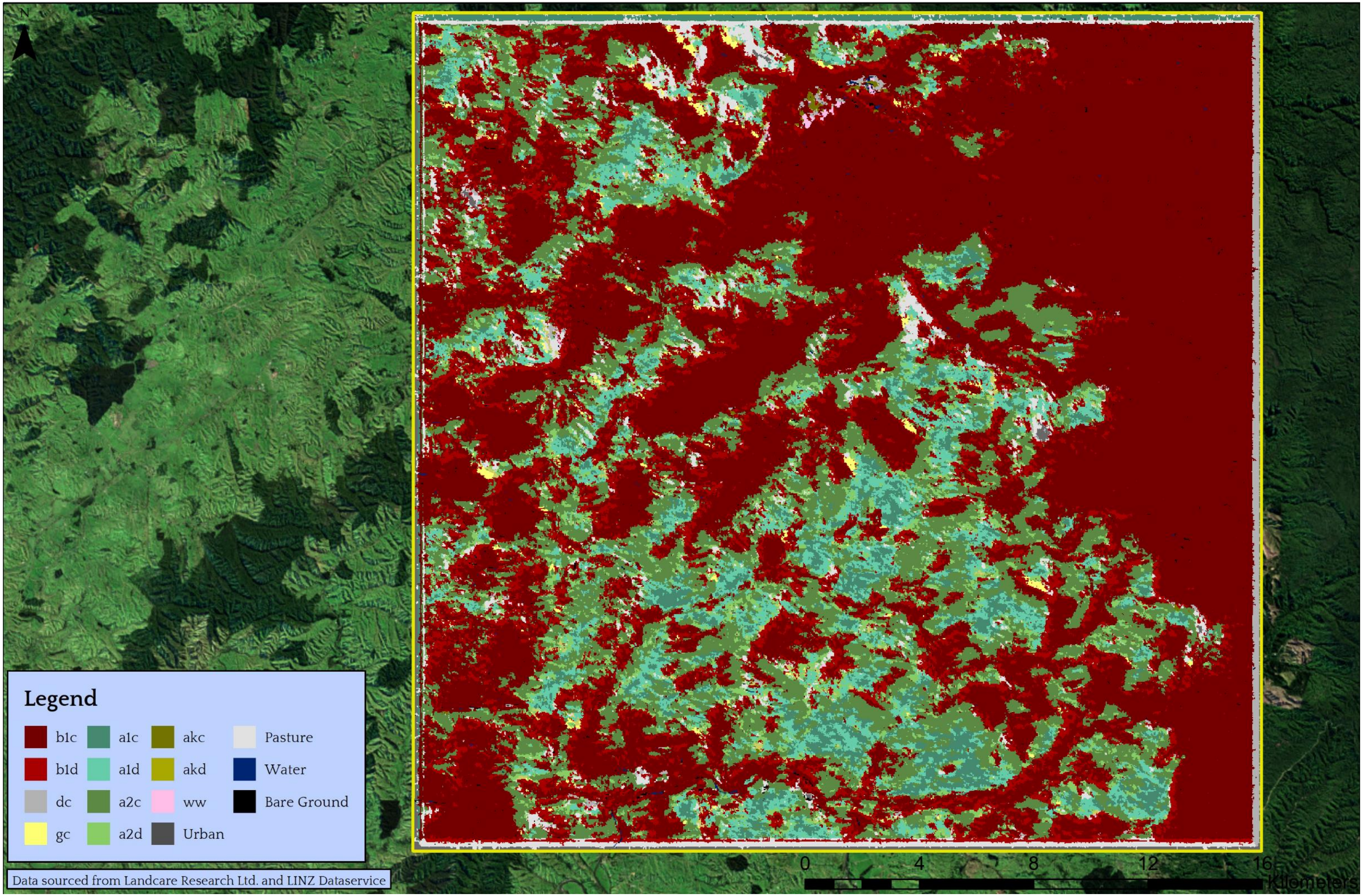


Table 8. Error Matrix of Manawatū-Whanganui AOI Classification (2016) for Accuracy Assessment

User Class \ Sample	b1c	b1d	Pasture	dc	akd	gc	a2d	a2c	akc	a1d	a1c	ww	urban	Sum
b1c	281	24	0	0	1	1	2	3	0	0	0	0	0	312
b1d	78	40	3	4	1	1	10	3	0	0	1	1	1	143
Pasture	2	3	183	6	2	3	50	11	0	10	3	0	8	281
dc	10	7	3	43	1	2	33	4	2	0	3	1	1	110
akd	2	0	5	1	30	9	15	4	25	3	6	2	0	102
gc	2	1	1	0	4	51	30	0	2	1	0	0	0	92
a2d	2	3	16	4	3	11	148	61	3	10	25	0	0	286
a2c	8	0	8	1	0	1	44	309	13	4	54	0	0	442
akc	4	0	1	3	9	4	9	13	65	0	19	0	0	127
a1d	2	2	12	3	0	1	75	42	4	87	215	0	2	445
a1c	0	0	0	0	0	3	20	35	1	33	772	0	0	864
ww	2	1	2	10	0	3	14	3	2	1	0	24	0	62
urban	0	0	2	2	0	0	0	0	0	0	0	0	287	291
unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sum	393	81	236	77	51	90	450	488	117	149	1098	28	299	
Producer	0.72	0.49	0.78	0.56	0.59	0.57	0.33	0.63	0.56	0.58	0.70	0.86	0.96	
User	0.90	0.28	0.65	0.39	0.29	0.55	0.52	0.70	0.51	0.20	0.89	0.39	0.99	
Hellden	0.80	0.36	0.71	0.46	0.39	0.56	0.40	0.66	0.53	0.29	0.79	0.53	0.97	
Short	0.66	0.22	0.55	0.30	0.24	0.39	0.25	0.50	0.36	0.17	0.65	0.36	0.95	
KIA Per Class	0.69	0.47	0.76	0.54	0.58	0.56	0.27	0.58	0.54	0.52	0.61	0.85	0.96	
Overall Accuracy	0.65													
KIA	0.60													
Sample Area (ha)	683.78	43.47	247.09	37.79	39.07	69.35	260.42	560.74	163.05	92.40	1550.56	18.25	233.09	
Sample Area as % of AOI	0.77	0.05	0.28	0.04	0.04	0.08	0.29	0.63	0.18	0.10	1.75	0.02	0.26	

Figure 30. Producer and User Accuracy Spread with KIA of Each Class (Manawatū-Whanganui AOI)

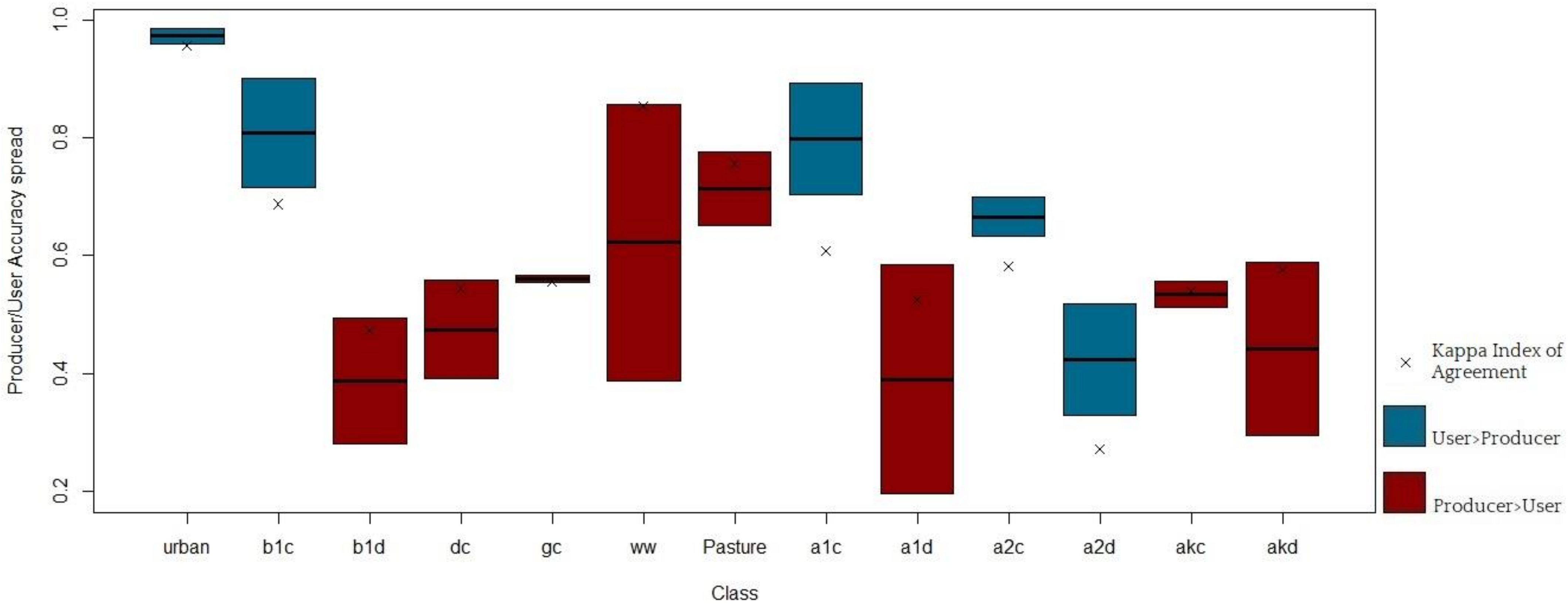
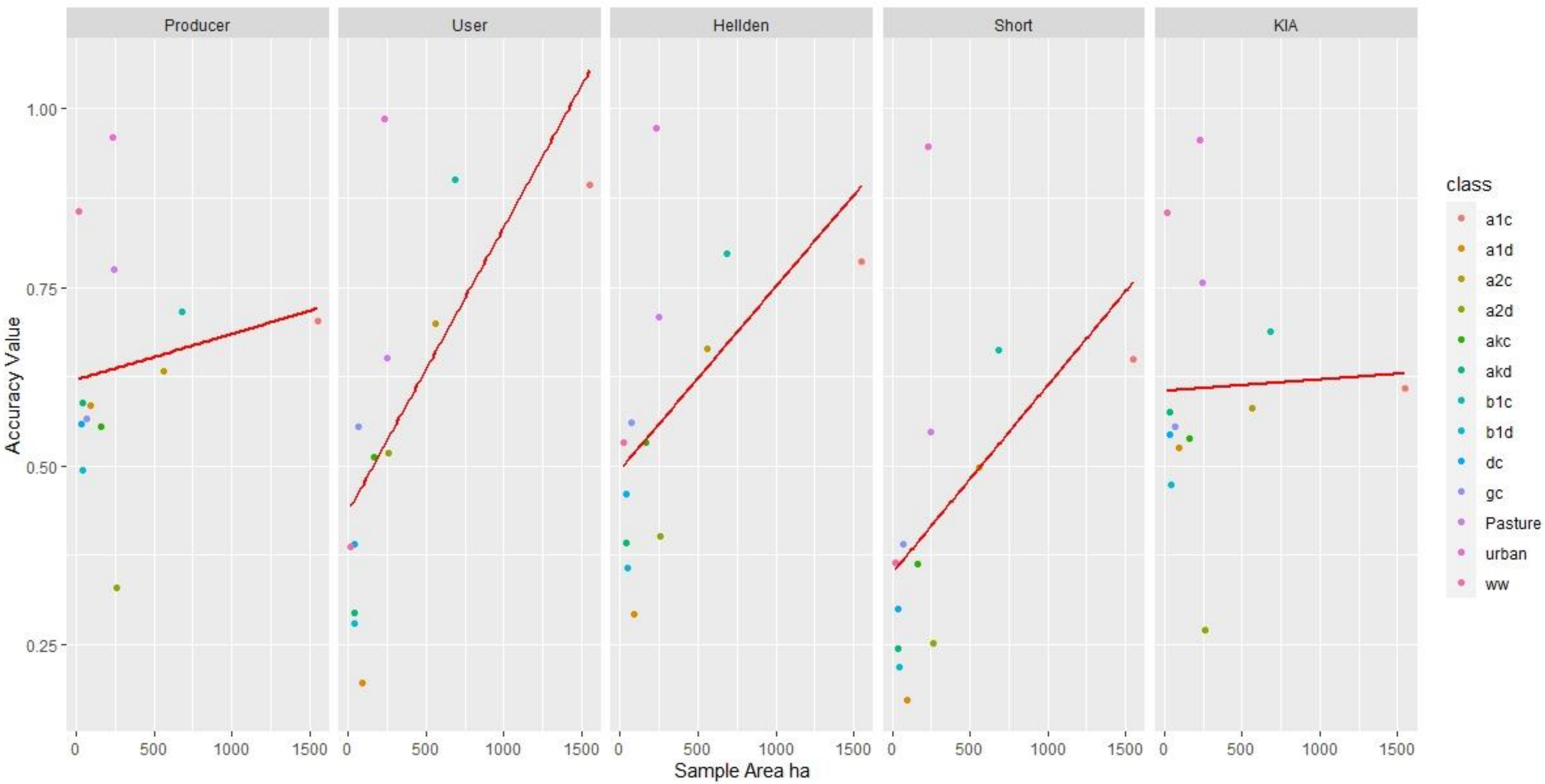


Figure 31. Simple Linear Regression Comparing Class Accuracy Measures With Their Absolute Sample Areas (Manawatū-Whanganui AOI)



4.7 Analysis of Automated Classification (Manawatū-Whanganui AOI)

Imagery

Spectral artefacts (from topographic correction) visible on large patches of dark bush in 2016 image.

Apparent harsh contrast in 1989 image.

Data noise visible over areas of uniform colour in 1989 image.

Classification Framework

The framework used in classification of this AOI included the eleven core classes but did not include continuous native broadleaved shrubland and diffuse native broadleaved shrubland classes. The classes exotic woody shrubland and continuous urban (the town of Taumarunui) are unique to this AOI.

Land Cover Change

The classifier, trained on primary satellite imagery, was unable to recognise the land cover classes from the final classification framework when applied to the secondary satellite imagery. Estimates of land cover change are of little use as it is obvious on visual inspection of the 1990 land cover map (Figure 29, p. 71) that the accuracy of the predicted data is very low. No reliable changes in land cover are observable in these data.

Accuracy of 2014 classification

Accuracy varies widely between classes but those defined as continuous in the classification framework were generally more accurately classified than those defined as diffuse.

The four classes that were consistently the most accurately classified across all metrics are: continuous urban, continuous planted exotic, continuous pasture and continuous native remnant.

The four classes that were consistently the least accurately classified across all accuracy metrics are: diffuse native remnant, diffuse native regenerating, diffuse kanuka/manuka and diffuse planted exotic.

As overall accuracy is 0.65 and KIA is 0.60, accuracy should be considered moderate (Congalton and Green, 1999) and it should be assumed that there are still significant errors in the methods used to produce these maps.

Effect of Ground Truth Sample Size on Classification Accuracy

A linear model was fit to each of the five classification accuracy measures to test the effect of sample size on classification accuracy (figure 31, p. 74). Each regression showed a general positive trend with high indicators of significance. R^2 values were poor however so all accuracy assessment values were pooled to test the overall relationship (n=65). This model's fit was also poor (R^2=0.1514).

These simple linear models do not explain enough of the variance in the effect of ground truth sample area on classification accuracy to be able to have confidence in the relationship demonstrated in the regression lines in Figure 31. The high residual standard error and low R^2 values indicate that there are likely other undescribed variables that strongly affect the accuracy of this classifier.

Figure 32. Corrected Primary Satellite Dataset (Northland AOI, Landsat 8, 2016) Superimposed Over Aerial Image Composite

Figure 33. Corrected Secondary Satellite Dataset (Manawatū-Whanganui AOI, Landsat 4, 1989) Superimposed Over Aerial Image Composite

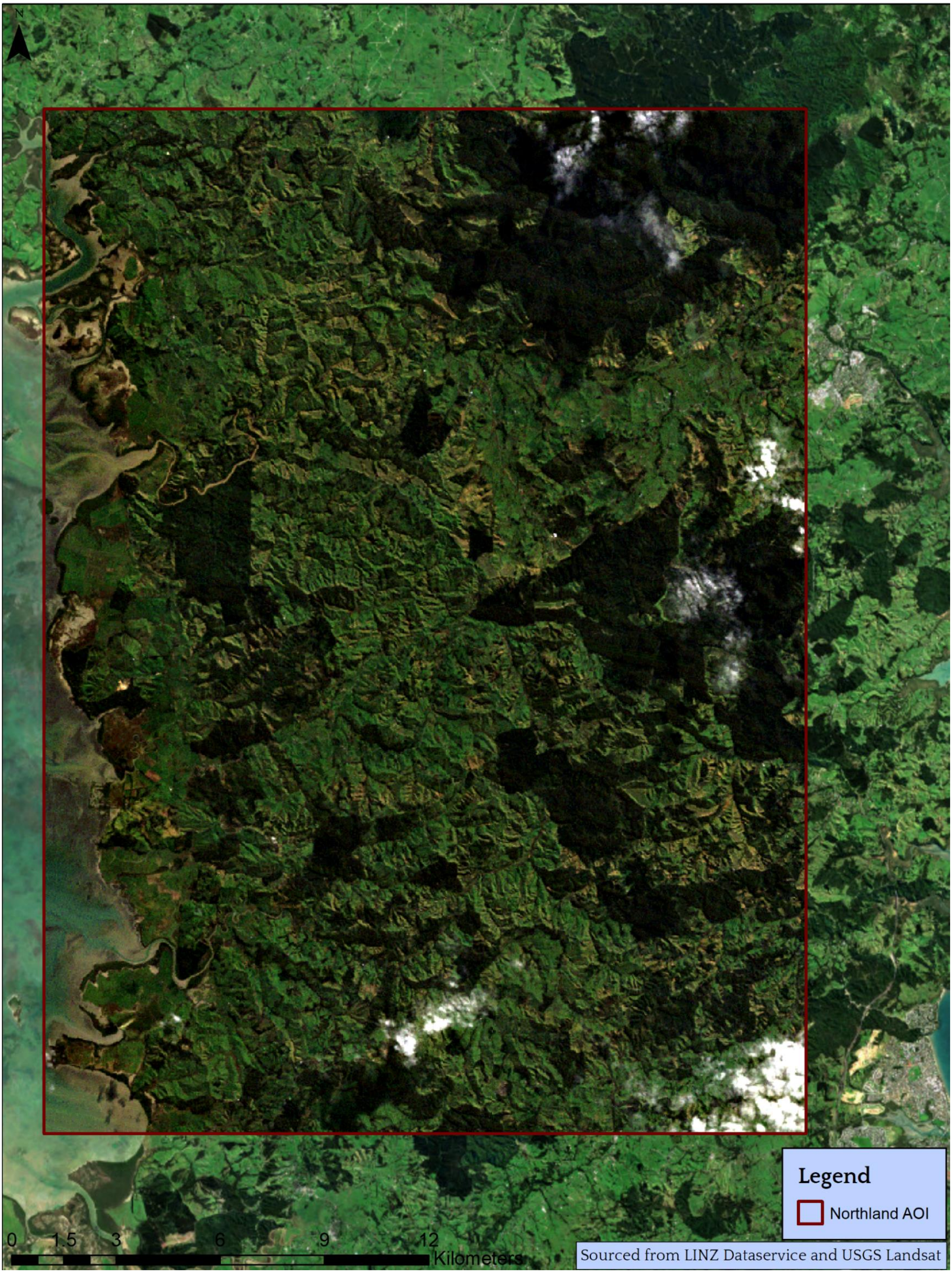


Figure 34. Automatically Classified Land Cover Map of Northland AOI (2016)

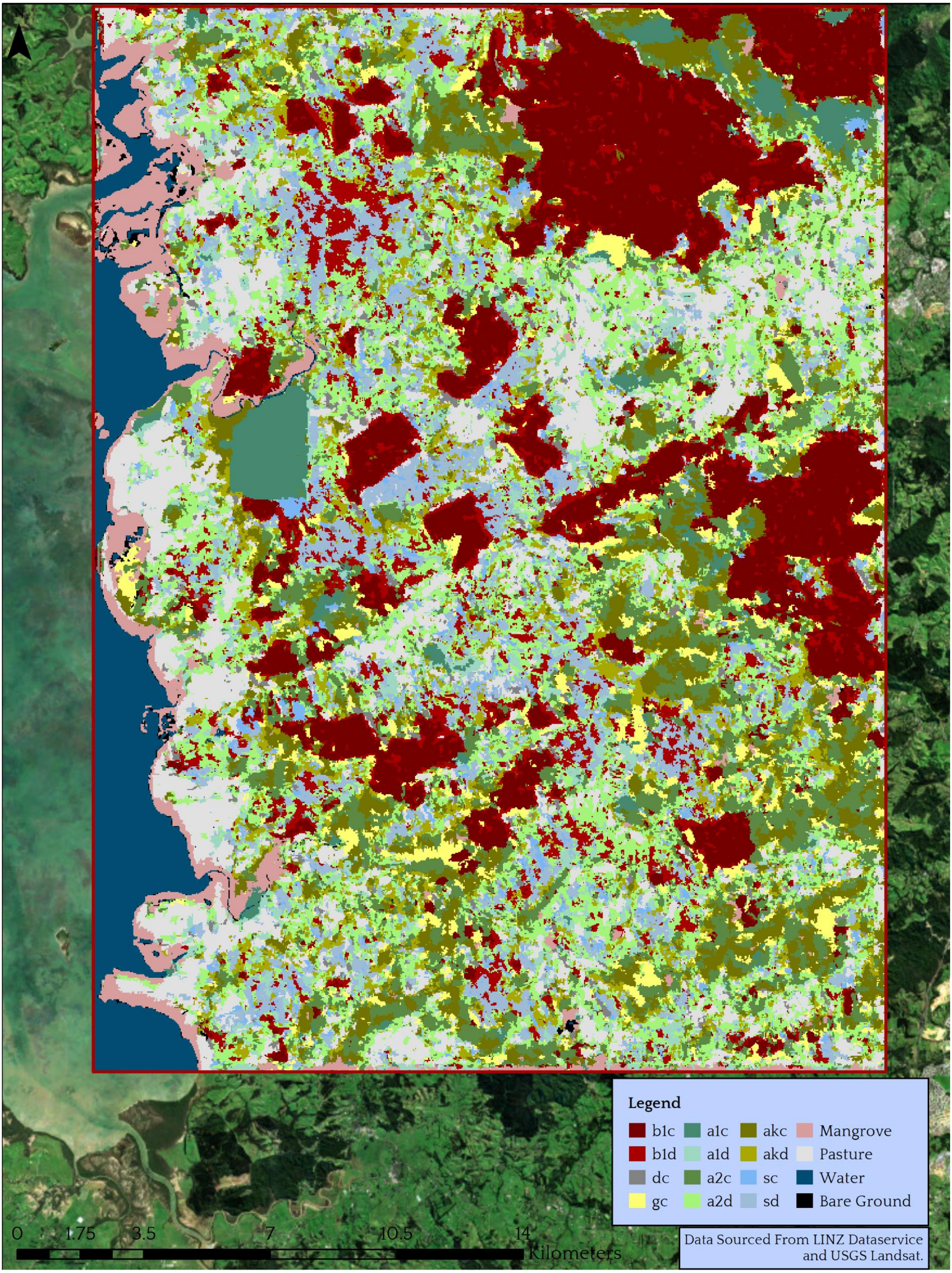


Figure 35. Automatically Classified Land Cover Map of Northland AOI (1989)

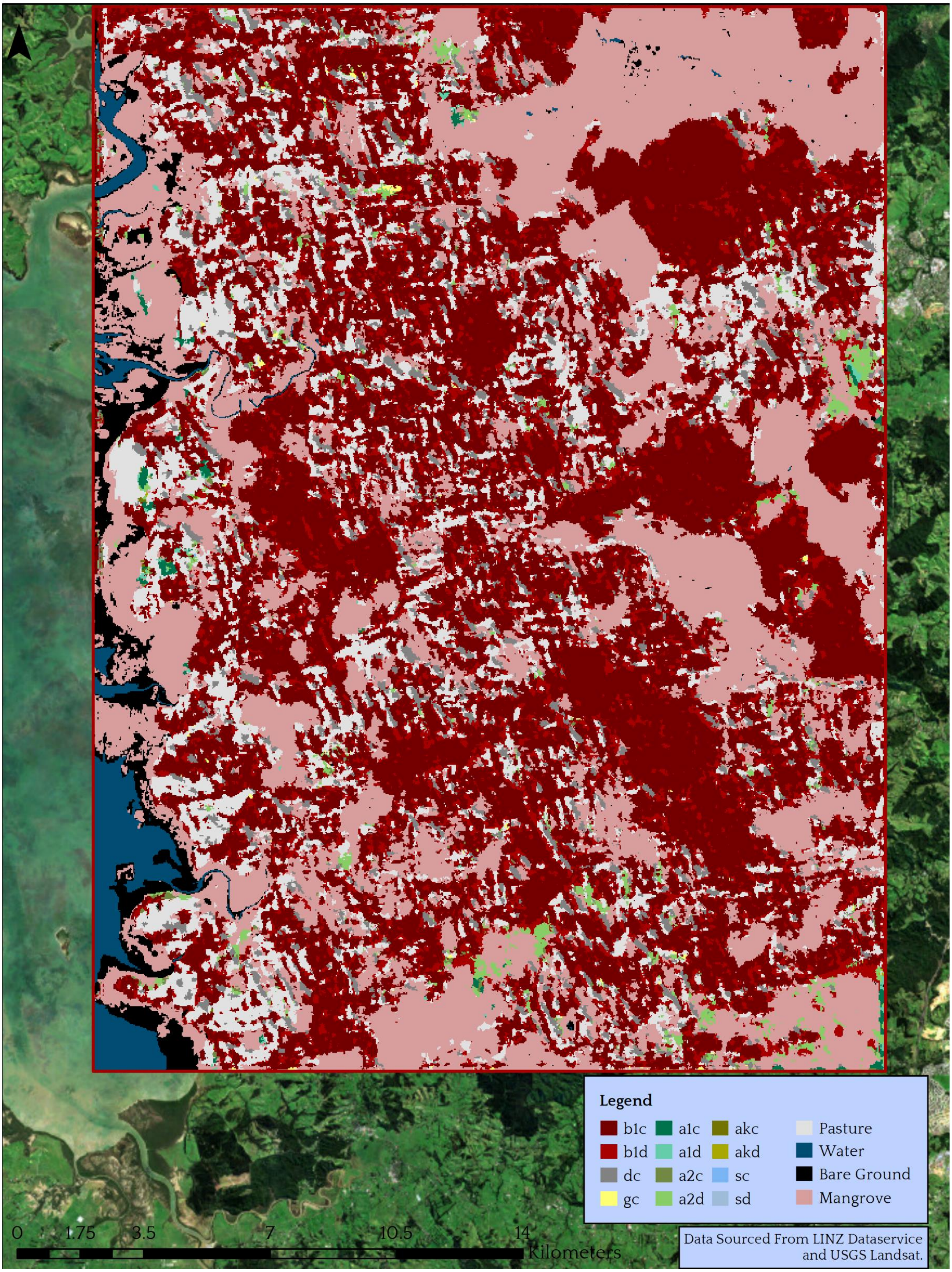


Table 9. Error Matrix of Northland AOI Classification (2016) for Accuracy Assessment

User Class \ Sample	b1c	b1d	Pasture	dc	akd	gc	a2d	sd	a2c	sc	akc	a1d	a1c	mangrove	Sum
b1c	356	76	0	0	2	4	2	0	10	0	6	3	0	1	460
b1d	84	99	4	0	2	2	14	4	16	0	4	2	1	0	232
Pasture	3	2	200	0	11	1	31	4	7	0	6	1	1	0	267
dc	0	0	3	6	1	1	8	1	1	1	0	0	0	0	22
akd	1	1	6	0	59	8	24	4	20	0	17	1	1	0	142
gc	6	8	3	2	4	105	6	3	9	0	2	0	1	0	149
a2d	5	11	19	2	14	17	163	5	116	0	6	1	3	0	362
sd	1	14	26	0	9	2	13	47	9	1	2	0	0	0	124
a2c	17	3	4	0	12	9	61	1	226	0	14	0	6	0	353
sc	1	2	11	2	3	3	6	18	2	15	1	2	2	0	68
akc	1	0	2	0	42	5	8	1	15	0	74	2	10	2	162
a1d	1	1	6	0	2	2	14	2	6	0	3	28	41	0	106
a1c	0	0	0	0	2	0	2	0	1	0	1	3	273	0	282
mangrove	0	0	1	0	1	0	0	0	0	0	2	0	0	42	46
unclassified	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sum	476	217	285	12	164	159	352	90	438	17	138	43	339	45	
Producer	0.75	0.46	0.70	0.50	0.36	0.66	0.46	0.52	0.52	0.88	0.54	0.65	0.81	0.93	
User	0.77	0.43	0.75	0.27	0.42	0.70	0.45	0.38	0.64	0.22	0.46	0.26	0.97	0.91	
Hellden	0.76	0.44	0.72	0.35	0.39	0.68	0.46	0.44	0.57	0.35	0.49	0.38	0.88	0.92	
Short	0.61	0.28	0.57	0.21	0.24	0.52	0.30	0.28	0.40	0.21	0.33	0.23	0.78	0.86	
KIA Per Class	0.70	0.41	0.67	0.50	0.33	0.64	0.38	0.50	0.45	0.88	0.51	0.64	0.78	0.93	
Overall Accuracy	0.61														
KIA	0.56														
Sample Area (ha)	767.73	108.26	389.36	5.11	97.26	114.22	184.91	62.19	460.67	22.33	149.70	22.96	709.02	81.77	
Sample Area as % of AOI	1.19	0.17	0.60	0.01	0.15	0.18	0.29	0.10	0.71	0.03	0.23	0.04	1.10	0.13	

Figure 36. Producer and User Accuracy Spread with KIA of Each Class (Northland AOI)

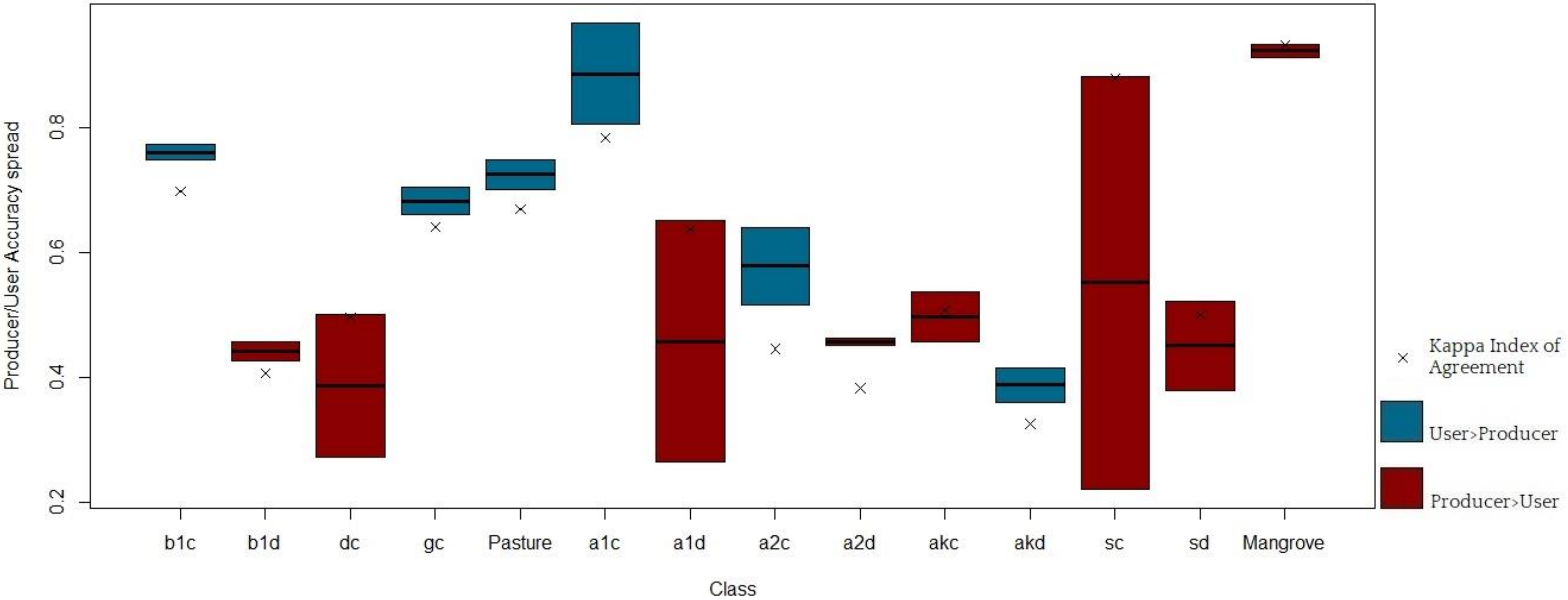
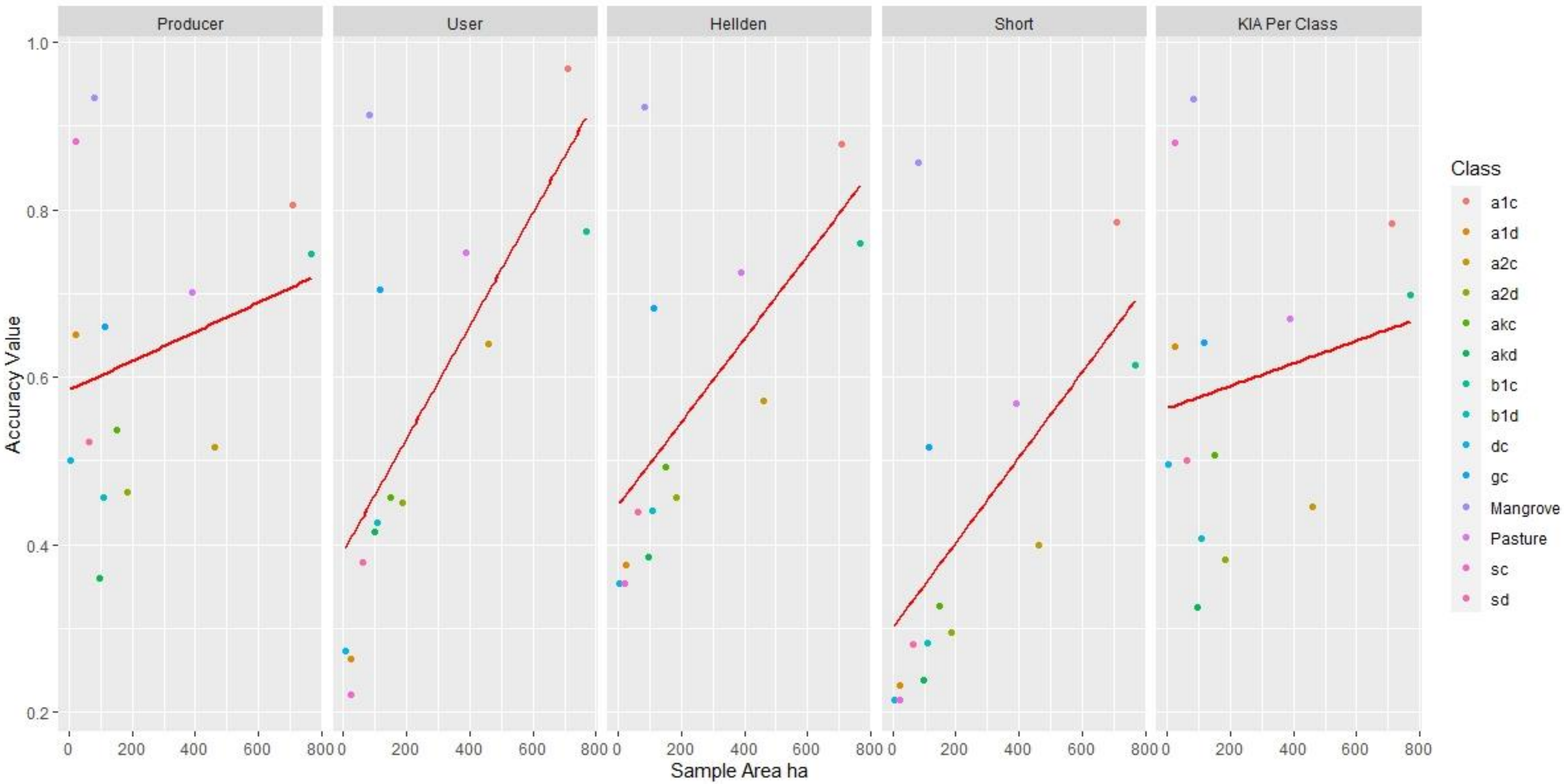


Figure 37. Simple Linear Regression Comparing Class Accuracy Measures With Their Absolute Sample Areas (Northland AOI)



4.8 Analysis of Automated Classification (Northland AOI)

Imagery

Spectral artefacts (from topographic correction) visible on dark bush areas in 2016 image.

Apparent harsh contrast and striated data noise visible over areas of uniform colour (sea) in 1989 image.

Of the three regions studied, the Northland AOI has the most complex arrangement of land cover classes and steep topography and is most likely to be obscured by cloud in the satellite data record. This is likely to have contributed to the poorer performance in classification.

Classification Framework

The framework used in classification of this AOI included the eleven core classes as well as continuous native shrubland, diffuse native shrubland and continuous mangrove (which was unique to this AOI). Fourteen model classes were used to classify this AOI while only thirteen were used to classify the previous AOIs.

Land Cover Change

The classifier, trained on primary satellite imagery, was unable to recognise the land cover classes from the final classification framework when applied to the secondary satellite imagery. Estimates of land cover change are of little use as it is obvious on visual inspection of the 1990 land cover map (Figure 35, p. 80) that the accuracy of the predicted data is very low. No reliable changes in land cover are observable in these data.

Accuracy of 2014 classification

Accuracy varies widely between classes but those defined as continuous in the classification framework were more accurately classified than those defined as diffuse.

The four classes that were consistently the most accurately classified across all metrics are: continuous mangrove, continuous native remnant, continuous planted exotic and continuous pasture.

The four classes that were consistently the least accurately classified across all accuracy metrics are: continuous broadleaved shrubland, diffuse broadleaved shrubland, diffuse kanuka/manuka and diffuse native remnant.

As overall accuracy is 0.61 and KIA is 0.56, accuracy should be considered moderate (Congalton and Green, 1999) and it should be assumed that there are still significant errors in the methods used to produce these maps.

Effect of Ground Truth Sample Size on Classification Accuracy

A linear model was fit to each of the five classification accuracy measures to test the effect of sample size on classification accuracy (Figure 37, p. 83). Each regression showed a general positive trend with high indicators of significance. R^2 values were poor however so all accuracy assessment values were pooled to test the overall relationship (n=70). This model's fit was also poor (R^2=0.215).

These simple linear models do not explain enough of the variance in the effect of ground truth sample area on classification accuracy to be able to have confidence in the relationship demonstrated in the regression lines in Figure 37. The high residual standard error and low R^2 values indicate that there are likely other undescribed variables that strongly affect the accuracy of this classifier

4.9 Overall Comments on Trends across all AOIs

The method developed in this research was unable to produce a time series of classified maps showing land cover change with any useful accuracy. However, analysis of the results that were attained from classification of the primary satellite datasets can provide direction for improving aspects of the method for future research. The most relevant observations from these results are stated below.

Sources of Classification Accuracy and Error

The wide range in classification accuracy between classes (Tables 7,8,9 and Figures 24,30,36) provides a good indication of which classes were easiest and most difficult for the classifier to identify correctly. As the classifier was able to predict the membership of pixels to some classes with high accuracy but was only able to achieve low or moderate accuracy in predicting others, it is likely that the classifier model is not the sole source of error. The effect of ground truth sample area on classification accuracy was tested as a way of explaining some of this error but the results were inconclusive, and no significant relationship was found that explained the differences in classification accuracy. The factor that was found to have the greatest effect on classification accuracy was the class description in the classification framework. As this controlled the contents of each ground truth sample it is likely that aspects of the class descriptions and ground truth sample quality are responsible for a large portion of the variance in between-class classification accuracy.

The comparison of user and producer accuracies can provide some insight into the ways in which reported values are inaccurate and can be used to assist decision making in changing class definitions and ground truth sample contents. Where producer accuracy is greater than user accuracy the spread suggests this class may be over-classified. Where user accuracy is greater than producer accuracy the spread suggests this class may be under-classified. The larger the spread between the two, the less confidence we should have in any reported measure of accuracy. This simple visualisation does not take factors of human error into account such as bias introduced during ground truth sample collection or digitisation. The unambiguous ground truth data production method and consistency of the analyst's decision making was used to control this source of error so that the user/producer accuracy comparison could be used to alter definitions in the classification framework with confidence.

Performance of Defined Classes

The acceptability and stability thresholds used to examine class accuracy measures in this section are entirely subjective and were defined in response to the class accuracy results. Future studies should define these thresholds prior to the production and analysis of results to ensure that the analysis framework is underpinned by objective measures of success.

If acceptable accuracy for a class is defined in this thesis as having both producer and user accuracies above 70% (constraining both accuracy and spread) then the classes that achieved acceptable accuracy are as follows:

Canterbury

- Continuous planted exotic
- Continuous pasture
- Continuous deciduous
- Continuous gorse
- Continuous kanuka/manuka.

Manawatū/Whanganui

- Continuous planted exotic
- Continuous native remnant
- Continuous urban.

Northland

- Continuous planted exotic
- Continuous native remnant
- Continuous pasture
- Continuous gorse
- Continuous mangrove.

Class accuracies were also assessed for stability across all three study sites. A provisional measure of stability was defined as: a difference of less than 0.25 between producer and user accuracies across each of the three AOIs.

The most stable classes present in all three classified maps were:

- continuous planted exotic
- Continuous native remnant
- Continuous gorse
- Continuous native regenerating
- Diffuse native remnant.

Most Accurate and Stable Classes

The classes that best satisfied the acceptability and stability criteria across all three AOIs were:

- Continuous Planted Exotic
- Continuous Native Remnant
- Continuous Gorse

Accuracy measures of these classes are compared with features of their class definitions and ground truth samples below.

Continuous planted exotic is the only class that was determined to have acceptable classification accuracy and stability across all three AOIs. There was little variation in KIA across the three AOIs and the mean of all three (0.7) indicates there is moderate agreement between predicted classification and validation samples.

Continuous native remnant and continuous gorse were among the most stable classes while also achieving acceptable accuracy in two of the three AOIs. KIA values indicate moderate agreement in each of the three observations of continuous native remnant with the highest value (0.78) in the Northland AOI where this class's largest ground truth dataset and best classification accuracy were also measured. KIA values of continuous gorse indicate strong agreement in the Canterbury AOI (where classification accuracy was highest but ground truth sample size was smallest) and moderate agreement when averaged across the three AOIs.

The continuous density classes described above were easier to sample than diffuse density classes. Large patches of continuous classes were often identifiable in both primary and secondary ground truth datasets and as the class definition requires each pixel to be mostly covered by one single class of land cover, ambiguity of a pixel's class membership was rarely encountered when drawing ground truth data polygons.

The continuous planted exotic and continuous native remnant classes had the largest ground truth sample areas in all three AOIs with the exception of continuous native remnant in Canterbury which had a relatively very small sample size. The Canterbury AOI was also the only place where the accuracy of continuous native remnant was below the provisional level of acceptability. The continuous gorse class had small sample size in all AOIs and did not pass the acceptability threshold of accuracy in the Manawatū-Whanganui AOI.

The measured classification accuracy of continuous planted exotic is the most reliable of any class based on stability and KIA values. The measured classification accuracy of continuous native remnant is also relatively reliable and the large sample sizes of these classes provides weak support for the hypothesis that accuracy increases as sample size increases.

Despite the assessment of good stability, low spread and good classification accuracy in two of the three AOIs, the reliability of the measured accuracy of continuous gorse is questionable. The high KIA value and low spread of continuous gorse in the Canterbury AOI may indicate that this ground truth sample pool was higher quality than its counterparts in the other AOIs.

Mixed Results Classes

The accuracy of all other classes can be considered either too low or too variable to be acceptable in this study. Notable features of class definition and ground truth samples of classes with mixed performance are described below.

The classification accuracies of continuous regenerating native and continuous native remnant were stable across all three AOIs but were too low to be considered acceptable in any of them. KIA values for both classes indicated moderate agreement, this value had low variation across the three observations. The diversity of NWV species potentially present in both classes is high and consequently these classes' ground truth data is likely to be highly spectrally variable.

Continuous regenerating native had a large sample size in Northland and Manawatū-Whanganui but the classifier was still able to achieve comparable accuracy in Canterbury despite the much lower sample size. This may indicate that the class definitions are higher quality in the Canterbury AOI, but equally it may indicate that this class definition is truly indiscriminable in these data and unsuitable for use in this classification process.

Diffuse native remnant had a low sample size in all AOIs, as this is a diffuse class the risk of including ambiguous or erroneous pixels into the ground truth sample is high. It is unsurprising that this class was not classified with high accuracy, however the consistency of these results does indicate that classification accuracy of this class may be able to be improved simply by increasing sample size.

The class pasture attained acceptable accuracy in both Canterbury and Northland but did not meet stability criteria. Sample size was large in all three AOIs, and higher accuracy was achieved where there were larger samples. KIA was highest in the Canterbury AOI where the sample size was largest but lowest in the Northland AOI which had a larger sample size than the Manawatū-Whanganui AOI. Additionally, relatively high spectral variation can be seen in pasture samples. These results may indicate that pasture may require a larger sample size than others to attain acceptable classification accuracy.

The ground truth samples of continuous kanuka/manuka were moderately sized in all three AOIs but the highest accuracy and strong agreement in KIA (0.88) was achieved in Canterbury where the smallest sample size was taken which opposes the hypothesis that increased accuracy is correlated with increased sample size. As uniform, continuous kanuka/manuka scrub was subjectively more difficult for the researcher to identify in aerial imagery in the Manawatū-Whanganui and Northland AOIs, inaccurate sample polygon drawing is likely to have been a significant source of error in these samples. Site familiarity is also likely to have played a role in the accuracy achieved by this class in the Canterbury AOI.

Diffuse regenerating native had a small sample size in Canterbury and moderate sample sizes in Manawatū-Whanganui and Northland. Despite this the best accuracy and KIA (0.55) values were recorded from the Canterbury site, again counter to the hypothesis that larger sample size improves accuracy. This may be partially explained by the fact that diffuse regenerating native was hard to accurately identify from aerial imagery and as it is a diffuse density class there was a much higher chance of encountering ambiguous pixels in the ground truth data. This result may also be partially explained by researcher's familiarity with the Canterbury site. This result (and that of continuous kanuka/manuka in the previous paragraph) suggests that sources of error other than sample size have a strong effect on some classes. Furthermore, the continuous and diffuse regenerating native class definitions have high species diversity, and ground truth sample pools that represent the resulting spectral diversity fairly may be difficult to produce. Analysis of the spectral variance in these classes' ground truth data samples may allow better class definitions to be written to lump or split classes into more spectrally separable sub-classes.

Two classes were present in only one of the AOIs: Continuous mangrove in Northland and urban in Manawatū/Whanganui. The data available for analysis of these classes' accuracy is limited as is the significance of any inferences made in the following two paragraphs.

The continuous mangrove class's sample size was small, and the data appears relatively spectrally uniform. Accuracy measures were high and classified areas appear accurate across most of the Northland AOI except for spotted areas in the south and west. The KIA was very high (0.93) indicating strong agreement. This class's CNN heatmap could have been manipulated prior to application of the RT classifier to limit continuous mangrove classification to coastal areas. This would have solved this specific misclassification but a more robust, classifier or ground truth level solution should be sought.

Although urban was accurately classified in the Whanganui/Manawatū AOI there was one relatively large and conspicuous region of misclassification in the far west of the map. Sample size was moderate to large, measured accuracy was high as was KIA value (0.96).

This class appears highly spectrally variable, analysis of this variation and adjustment of ground truth data selection methods may provide solutions to reducing classification artefacts.

Least Accurate and Stable Classes

The remaining classes all had small sample sizes, poor classification accuracy and poor stability. Analysis of spectral variation may be helpful in developing methods of improving class accuracy. Classes that were not acceptably accurate or stable in any of the three AOIs were:

- Continuous deciduous
- Diffuse planted exotic
- Diffuse broadleaved shrubland
- Continuous broadleaved shrubland
- Diffuse kanuka/manuka

General Trends in Classification Accuracy

Continuous classes were classified more accurately than diffuse classes across all three AOIs. Classification accuracy of these classes was improved through repeated targeted refinement of ground truth samples in response to the classification accuracy metrics produced in eCognition. Continuous native broadleaved shrubland was the only continuous cover class that was judged to be relatively poorly classified. This is likely the result of the high ambiguity of shrubland pixels in the aerial imagery causing inconsistency in ground truth sample production

The linear regression used to model the relationship between class ground truth area and classification accuracy performed poorly. The sample size and accuracy measures of the most accurate and stable classes did support the hypothesis that larger sample sizes increase classification accuracy, but as the sample size of the between AOI comparisons is so low (n=3), a more thorough statistical analysis of the relationship between ground truth sample size and classification accuracy is not likely to produce useful results in this thesis.

Many of the classes estimated to have the most spectrally uniform ground truth datasets achieved acceptable classification accuracy and the classes estimated to have the most spectrally variable ground truth datasets were classified most poorly. This observation requires further statistical exploration as this study did not quantify or analyse the spectral variation of classes across each of the six bands used for classification.

Overall, the best results were achieved in classification of the Canterbury AOI despite having the smallest total area of ground truth samples. This indicates that regardless of the existence of a correlation between classification accuracy and ground truth sample area, sample quality is likely to have a stronger influence on a class's classification accuracy. The Canterbury AOI was well known to the researcher and it is assumed that this improved the quality of field sampling and aerial photography interpretation. Repeat visits to ground truth sites were helpful for confirming drawn samples and the estimating the extent of patches of classes that are cryptic in the secondary ground truth data (e.g. continuous and diffuse broadleaved shrubland) could be better estimated when the researcher was very familiar with a property within the AOI. Results of this analysis of class accuracy indicates that improvements in ground truth data quality are more likely to produce a significant increase in classification accuracy than any other variable examined in this thesis.

5. Discussion

This section will detail the stages of research and development which produced each of the workflows described in the methods section. A large portion of the development phases of my research consisted of iterative testing of data processing and formatting methods to ensure the workflow was easy to replicate. This was both for the benefit of future researchers and so that the effect of altering key parameters of any of the workflows could be examined over the full model. Although the body of remote sensing literature contains a wide variety of research which was used to direct the development of these workflows, the highly specific nature of using Landsat imagery to produce a time series of land cover change on New Zealand sheep and beef farms using a hybrid machine learning classifier meant that many of the finer details of these workflows were unprecedented.

This section is also intended to discuss the results of this research in contrast with the overall scope of the data available in the New Zealand LCDB. The LCDB is able to show attributes of land cover at five discrete time periods between 1996 and 2018 (examples are shown in Appendix F) and can discriminate a wide variety of land cover classes. However, it is unable to estimate land cover from any time period earlier than this, and does not account for the density or history (remnant/regenerating status) of native woody vegetation used in the final classification framework (Table 6, p. 54).

5.1 Satellite Data Selection

The Landsat programme was chosen as the data source for satellite imagery in this research due to its free availability, long period of operation and wide multi-band spectral range. Although other satellite imaging systems may fulfil some of these criteria or have unique innovations (e.g. SPOT, WorldView, Sentinel, Ikonos), the Landsat catalogue is unrivalled in its applicability to the goals of the current study.

Data used in the final analysis was selected from the Landsat Collection 1 Level 1 archive. The Earth Explorer data portal used in this research also provides access to Landsat Collection 2 Level 2 data which contains pre-corrected surface reflectance data products. These data were considered for use in this research but were discounted as the timeframe this collection covers is shorter than Landsat Collection 1 Level 1 and the corrections applied did not include topographic corrections. These factors made it more complicated to correct and less useful to the aim of building a time series of land covers.

Although the Landsat missions have been in operation since the early 1970s with a nominal return period of sixteen to eighteen days, good quality scenes that cover a project's AOI at the required point in time can still be scarce. Data gaps in the Landsat catalogue for the

AOIs used in this research were a significant issue. Gaps are more frequent in earlier datasets (Landsat 1-5) and often are the result of deliberate choices by the USGS to prioritise imaging of America. The cloudy nature of Aotearoa New Zealand's meteorological environment compounds the data gap problem as a large proportion of all the Landsat scenes available are significantly obscured, especially in the more northern and western regions of the country. These challenges made acquisition of data sets across the full time-period as originally planned impossible, resulting in the available time series reducing from approximately forty years to twenty-six and twenty-eight years.

The earliest good quality images in the Landsat Collection 1 Level 1 catalogue were acquired in 1989 and 1990. Despite these difficulties, the Landsat catalogue proved to be comprehensive enough to allow a decent collection of high optical quality scenes to be sourced. Satellite data selection was successful in that, images were able to be selected that showed land cover at a date earlier than the earliest available iteration of the LCDB in 1996 (Landcare Research, 2020).

Although relatively high quality scenes from 1989/1990 were available, these data had some notable and unavoidable limitations which lowered the potential classification accuracy. The secondary satellite images of all three AOIs suffered from general spectral noise. This is most visible on areas of uniform colour (e.g. sea, continuous forest) and may appear in a striated or random pattern. The secondary satellite image of Canterbury (Figure. 21, p. 60) had a significant area of land obscured by a spectral artefact that appears to be caused by a digital processing error. This went unnoticed during the correction and data processing phases of the work and only became obvious during application of trained classifiers as it is present only in the NIR band. Translucent cloud affected some small areas of both primary and secondary images of the Northland AOI (Figures 32, 32, pp. 77,78), altering digital number values of these areas. Despite these problems, the chosen images were high quality relative to the pool of suitable images found on Earth Explorer, the next four paragraphs describe the low quality data that was examined but omitted from analysis.

Landsat missions 1-3 were in operation between 1972 and 1983 and acquired image data with the Multi Spectral Scanner (MSS) sensor (USGS, 1997). Land cover data from this period would be extremely valuable in estimating the distribution and arrangement of land cover far earlier than the available LCDB data, but poor image availability, sensor comparability and quality issues have meant that data acquired by MSS sensors were omitted from further investigation in early stages of this research.

Imaging of New Zealand began with Landsat 2 in 1975 and even then, published imagery from Landsat 2 and 3 is scarce. Despite these satellites having a return period of 18 days, the frequency of low quality data acquisition was much higher in these older sensors and most of the imagery covering this study's AOIs from the 1970s and 1980s is omitted from the collection 1 Level 1 catalogue because of quality concerns. Low quality image data is filtered out by the USGS image processing laboratories and is not available for download as a data product through Earth Explorer. Spectral artefacts, geometric artefacts and very high cloud cover are the main reasons why a scene's raw data may not meet the criteria required to be processed into a publishable data product. These older sensors and their associated calibration systems (both on-board and on the ground) were more prone to producing data that fell outside of the USGS' standards for radiometric and geometric error. Over time as the Landsat program's methods have improved, their data products' error rate has decreased, fewer scenes fail to meet publication criteria and consequently there is more consistency in the image timeline (USGS. 2019).

Despite the paucity of data in the early 1980s a few published scenes with low cloud cover were found that were initially considered as candidates for analysis. These Landsat 2 MSS scenes were eventually excluded from analysis due to difficulty with atmospheric and topographic correction. The two commercial software suites trialled (ReSe ATCOR and PCI Geomatica) are unable to process MSS data, likely as a consequence of both software packages being based on MODTRAN. A customised method for applying equitable corrections to MSS data could be developed, but a significant investment would need to be made into developing a software process for application to the specific images that would allow compatibility with corrected imagery produced by later Landsat sensors. If this data was thought to hold valuable information for land managers in future, it may be a line of research that would be worth pursuing but at present, this data is unavailable for land cover change research.

Malfunction of the Landsat 7 sensor's Scan Line Corrector (SLC) occurred on the 31st of May 2003 and despite repair efforts its failure was permanent and will affect all imagery acquired until the satellite's planned decommissioning in late 2020 (USGS, 2020). The SLC corrected for the effect of the satellite's forward motion on the side to side scanning action of the 'whisk broom' (across track) sensor. Although 78% of the pixels in each affected scene appear as normal, the extensive and regular gaps in the data would have both, significantly increased time taken to process these scenes equitably and significantly decreased accuracy and confidence in the result if they were to be included in the analysis (USGS. n.d.). For these reasons, no ETM+ data was used in this research.

Although the Landsat data catalogue does provide imagery of New Zealand as far back as 1975, because of the issues discussed above it was not possible to analyse many of the early images as originally planned. The imagery from 1990/1989 will still allow production of a time series that describes land cover 6-7 years earlier than the first version of the LCDB and the results attained in my research indicate that the six Landsat bands (available in imagery produced by Landsat 4 and later) used have some potential for discriminating land cover density classes and remnant/regenerating forest status. The selected data are an excellent resource for historic land cover change, but the methods of measuring this need to be carefully developed to ensure that the most important attributes can be visualised with accuracy.

5.2 Ground Truth Sample Production

Development of Ground Truth Data Format

The most important outcome of this research was the identification of the need for high quality ground truth data and the development of a method to produce it. The (pixel-tracing) method of combining primary and secondary ground truth datasets into a single set of ground truth samples by tracing polygons around pixel edges (described in Section 3.2 p. 22) was reached through a series of iterative data format tests. Initially the quickest and lowest effort method was used to draw land cover class samples, and the combined ground truth dataset tested for its suitability as a training set for a simple classifier model. When classification did not achieve acceptable results a specific change to the method was made and the entire combined ground truth dataset redrawn. Each iteration of polygon drawing method increased in complexity and time consumption until a functional solution was found that allowed reasonable classification accuracy to be produced.

The aim of this process was to discover how to maximise the accuracy achieved by a classifier model while minimising time taken to prepare high quality ground truth data using the selected software. Modern machine learning classifiers are now better at discerning signal from noise in imagery than ever before and therefore are able to produce more accurate maps from lower quality training data. As the results of my research indicate that ground truth data quality strongly affects classification accuracy, it will be critically important to ensure that an acceptable balance between accuracy of automated classification and effort required to classify land be found. The approach taken to development of the pixel-tracing method therefore aimed to maximise the ability of the classifier to produce accurate land cover change estimates and minimise the time investment required to produce high quality ground truth data.

Details of the unsuccessful methods of producing the combined ground truth dataset and their points of failure are outlined briefly here. Reasoning for their failure is suggested, but this should be considered in context of the software and hardware used in this study. Some of the assumptions that allowed for classification failure in the first two iterations of the combined ground truth dataset may not be applicable in future as the ability of machine learning classifiers to automatically discern signal from noise improves.

This study made use of an Object Based Image Analysis (OBIA) method to classify satellite imagery. OBIA methods use image segmentation algorithms to group spatially adjacent and spectrally similar pixels in an image into objects for use as the smallest units of analysis. This is an alternative approach to pixel-based image analysis which uses the image's pixels themselves as the smallest unit of analysis.

Ground truth for training and validation of classifiers is often applied to the imagery as point data in OBIA studies (Phiri, D., personal communication). The first iteration of the combined ground truth data set was comprised of point data extracted from the primary ground truth dataset. This approach relied on the segmentation algorithm to define boundaries of the ground truth samples; where a point fell within an object, all the pixels contained were assigned to the point's class. This approach has been shown to be effective in research where the AOI was very large and the ratio of the pixel area to AOI area was small (Phiri et al., 2019; Vittek et al., 2014). In my research however, the segmentation algorithm did not define the edges of objects accurately enough to the edges of vegetation patches and the variation introduced by erroneously included pixels caused classification accuracy to suffer. After achieving only poor results with this method, trials with ground truth data in point format were abandoned and all subsequent methods defined ground truth samples with polygons.

The next iteration of the ground truth data production method used the primary ground truth dataset to draw polygons over the land cover visible in the secondary ground truth data with the aim of representing the boundaries of patches of vegetation as accurately as possible. This approach assumed that combination of eCognition's "Vector Layer Import", "Multiresolution Segmentation" and "Assign Class by Thematic Layer" algorithms would simply assign a class to an image object based on the polygon with the largest area of overlap. This assumption proved to be erroneous as this approach produced classified maps with very poor accuracy and when the results of this algorithm were examined, they were found to behave differently than anticipated in a multitude of ways. For example some image objects were classified from only a very small overlap with a class sample polygon or a class that did not hold the majority overlap of a pixel's area, while some were left unclassified even with a seemingly significant overlap with a class sample polygon.

Furthermore, this method used the target land cover classification framework (Defined in Appendix A) and assigned large areas of mixed pasture and low density woody vegetation to 'sparse' density classes with coarsely defined boundaries. Sparse density woody vegetation covers large areas of each AOI and is an important attribute of NWV patch continuity at a landscape scale in my study areas, so identification of these classes was a high priority goal of this research. This method of identifying sparse woody vegetation as a land cover class proved to be incompatible with the way in which eCognition's segmentation algorithm handled the Landsat rasters. The drawn sparse woody vegetation samples likely had too much overlap in definitional features (chosen by the classifier) with pasture and higher density woody vegetation classes and only very low accuracy classified maps could be produced. Failure to successfully train a classifier to detect sparse woody vegetation classes drove the revision of both the sample selection method and the land cover classification framework.

An assumption made early in the research period was that a small amount of incorrectly classified ground truth data would not problematically skew a ground truth dataset containing a sufficiently large number of pixels. On examination of results of early classification trials however, it became clear that it was critically important to draw accurate boundaries around land cover sample patches and to create this dataset at the same spatial resolution as the satellite imagery. Additionally, it was found that polygons drawn to fit natural land forms and patch edges were handled by eCognition's segmentation algorithm in an unexpected manner. This drove the development of more accurate pixel-tracing through use of a grid ("fishnet") aligned with pixel edges that polygon edges could be snapped to. Implementation of this method on the entire ground truth dataset of one AOI led to a significant increase in classification accuracy and this method was used to improve the ground truth datasets of both other AOIs as well. This pixel-tracing method greatly improved classification accuracy and supports the hypothesis that high quality ground truth data is the most important input factor influencing classification accuracy in this research.

As the lowest effort methods failed to produce maps with any accuracy, the scale of image objects examined was reduced so that polygon edges could be drawn with better accuracy. This required increased reliance on the secondary ground truth dataset to allow pixels to be examined individually and polygon edges to be drawn to more strictly include or exclude areas pixel by pixel. The classification framework defined continuous and diffuse classes by percent coverage thresholds, these could be more accurately estimated when pixels were examined individually. For example, a diffuse class under the revised sample selection method would be a pixel that contained 15-70% of a woody land cover class while the remaining area was covered by pasture or bare ground. A revised ground truth dataset was tested in the classifier, continuous and diffuse land cover classes showed modest accuracy, but sparse classes were still poorly classified and so these were removed from the classification framework entirely.

The pixel-tracing method of ground truth sample digitisation could be considered an object based method meaning it delineates objects (usually but not necessarily comprised of multiple pixels) from the data for use as the smallest unit of analysis in the classifier. The structure of the pixel-tracing method does however simulate some of the characteristics of a pixel-based method in its small spatial scale and requirement to adhere polygon edges to pixel edges. This mixed approach allowed the classifier to examine the spectral values of image objects as the mean of all contained pixels (post segmentation) while ensuring the carefully drawn external boundaries of objects were retained. It was found that keeping the object scale small allows for observation of much of the detail that can be attained from pixel based classifiers while reducing the chance of producing noisy classification results (i.e. the salt and pepper effect). Keeping the object scale small also allowed much of the area that would have been considered 'sparse' (under the original conception in the target classification framework) to be accounted for by the diffuse classes.

The aerial imagery was used to generalise the patches of land cover into the classes in the classification framework and determine accurate boundaries through estimation of percentage of area covered by pixel. The effect of this process was the resampling of the primary ground truth dataset to match the data specifications of the Landsat imagery with addition of spatial attributes derived from both the aerial and primary satellite imagery using the Landsat pixel as the smallest unit of analysis. Although the classifier used the image object as the smallest unit of analysis, the combined ground truth data was produced by using the pixel as the smallest unit of analysis to preserve detail.

Aerial imagery was heavily relied on for improvement of ground truth data and classified maps through pixel-tracing and the assumptions made about these images' comparability to the primary satellite data are the most important identified source of error in the combined ground truth dataset. The primary ground truth dataset was used to provide detailed descriptions of patches of land cover as they are today, but the aerial imagery was acquired much closer temporally to the primary satellite datasets than the primary ground truth data was. This temporal gap should be kept as small as possible to reduce the chances of introducing error into the ground truth dataset through accidental inclusion of areas of sudden land cover change e.g. plantation forest harvest areas, ploughed pasture, burnt vegetation.

Future research may look to ground truth sample optimisation as a method of improving the representativeness of the ground truth dataset. Large scale structured data collection and mining approaches such as those developed by Chang and Kak (2019) may allow optimisation of the problem of producing the highest quality ground truth data in a minimised time frame.

Classification of Cryptic Land Covers

The majority of automated land cover classification methods described in the literature extracted their ground truth data from the same dataset used later for classification (e.g. Phiri et.al., 2019, Sharma et.al., 2017, Song et.al. 2019). This study's primary satellite image was used to aid creation of each of the combined ground truth datasets, but its use was very limited and the key spatial and classification choices were made largely through examination of the primary and secondary ground truth datasets. Many of the classes defined in this project's classification framework (Table 6, p. 54) were not able to be reliably identified in the satellite image data by a human observer these were referred to as 'cryptic classes'. The low spatial resolution of the satellite data and highly specific classification framework meant that many classes in the satellite data were indistinguishable from at least one other class on visual examination. This meant that this research relied on three main datasets to achieve its goals: one medium resolution satellite image for automated classification; one manually collected polygon dataset for primary ground truth; and one high resolution aerial imagery dataset for supporting (secondary) ground truth. These data were combined manually using human discretion to estimate the class and percentage land cover of each pixel. Once the problem of cryptic classes was identified, the aim of the initial classification phase of this research changed to: investigate how the six spectral bands common to each Landsat scene can be used to distinguish the nominated land cover classes, including those that are cryptic to the human observer.

Most classes in the classification framework should be considered cryptic in the satellite data to some degree. The only classes that are usually not cryptic to the human observer are: continuous planted exotic, water, bare ground, and pasture. The difficulty in detecting spectral differences between many native forest classes is likely why the LCDB has never distinguished between native remnant and native regenerating forest classes. The detection of cryptic classes through the inference-based ground truthing method described previously was a novel and challenging part of this research as there is little precedent for this in the literature. Although mapping accuracy in all final products are below the eighty-five percent accuracy generally considered acceptable in remote sensing (Congalton & Green, 1999), the detection of cryptic classes was successful enough to provide proof of concept for two of the key aims of this study: the accurate discrimination of different density classes of woody vegetation and discrimination of remnant forest from regenerating forest.

Development of Classification Framework

Developing clear and specific class definitions for the ground truth sample selection process will be a critical task for moving this research forward. If class definitions are not appropriately detailed, the limitation of Landsat pixel size (30x30 metres) could cause large areas of land cover to be obscured when automated land cover classification is carried out on a landscape scale. For a land cover class definition to be appropriate to the objectives of this thesis it must fulfil two central criteria:

1. Describe land cover using parameters that allow the classifier to predict the class of an image object with minimal accuracy error (maximise significant statistical difference in spectral values and spatial attributes between classes).
2. Describe land cover using details relevant to the environmental attributes of functional forest type and spatial arrangement (i.e. vegetation density and history).

The approach taken to class definition development in this thesis was to start with a classification framework designed to fulfil the second criteria (Appendix A: Target Land Cover Classification Framework) and test the ability of a classifier to separate these classes. The pixel data pool of poorly classified classes was examined and if the data was highly variable the class was removed from the framework or the definition was changed to be more specific and detectable. This process required many cycles of iterative testing of the classifier to reach a classification framework that was considered to be a suitable compromise between criteria 1 and 2 (Table 6, p. 54). The specific details of criteria 2 were adapted from Forbes et al. (2020) and were intended to ensure that the classification framework used in my research had a descriptive advantage over the classification framework used in production of the LCDB.

Better accuracy is likely to be achievable if this problem is approached by developing two frameworks to satisfy both criteria in parallel. Where a large framework of land cover classes is defined from ground truth sample statistics and ranked by significance of statistical difference. These classes are then assigned to the predetermined framework of classes relevant to environmental attributes for presentation so that these classes can be predicted and mapped with optimal accuracy. This method of assigning a separately determined land cover class to a group of classes which are defined purely by the statistics of spectral and spatial attributes would be more similar to the method used by a human classifying land cover manually through visual examination.

Ambiguous pixels were a common problem found in the ground truth digitisation process. Pixels often contained regions of ambiguous land cover where the combination of primary and secondary ground truth datasets did not allow the researcher to have high confidence in their class designation. While the most obviously ambiguous pixels were omitted from the combined ground truth dataset, many were undoubtedly included where ambiguity was not identified or some land covers were not visible in the (RGB) secondary ground truth data layer. Furthermore, a large number of pixels were not observed directly in primary ground truth data collection and their class was estimated (e.g. the interiors and back edges of large patches of land cover). Even where confidence in a pixel's class membership was high, there was still a significant risk of misclassification when producing the combined ground truth dataset. Each ambiguous pixel that was misidentified and labelled as an incorrect class allowed the classifier to widen the statistical definition of this class within the internally selected features. The more this misclassification error is repeated in the ground truth digitisation process, the more biased the classifier will become towards obscuring any class that contributed to that sample's mean spectral values that is not accounted for in its class definitions. This could become a significant source of error in the classifier model and cause large scale inaccuracy in the produced classified maps.

Each definition within the final classification framework used in this research (Table 6, p.54) is a simplification of the actual land cover found within a given Landsat pixel. The written class definitions in combination with the data's spatial resolution is a key limitation in the descriptive power of the produced classified maps. The best example for illustration of this is the problem of low density land cover classes. In this research the definitions of diffuse classes only allowed them to be assigned to a ground truth pixel where select woody vegetation classes occurred in combination with pasture or bare ground classes. The consequence of this definition constraint was that after the classifier was applied, ambiguous pixels were considered to be uniform in the analysis and that area of mixed land cover was obscured. This loss of area is very small at the single pixel or image object level, but at a landscape level (tens or hundreds of thousands of hectares) it is likely to significantly reduce accuracy in the final map product.

It is likely that the classified maps of primary satellite data produced in my research obscured substantial areas of woody vegetation classes where mixed class pixels were designated a class definition that did not fully describe their contents. The mechanism of the developed hybrid machine learning classifier provides a novel opportunity to assign a more complex and descriptive classification framework of functional land covers to Landsat data. The CNN algorithm (new to eCognition 9.5 in May 2019) produces a probability heatmap for each model class rather than directly classifying image objects, future research should consider developing a method of inferring density classes from heatmaps produced through initial classification of continuous classes only. For example, under the current structure the land cover of a pixel with CNN heatmap probabilities of 0.5 for membership to continuous kanuka/manuka and 0.5 for membership to continuous planted exotic is likely to be covered by a mixture of both classes but will be classified as only one by the secondary (Random Trees) classifier. If model classes were strictly continuous and low density classes were determined through heatmap sampling or threshold classification however, this pixel would be able to be classified as a mixture of diffuse kanuka/manuka and diffuse planted exotic which would allow for greater accuracy in post classification area analysis.

The classification framework and land cover area calculations would be somewhat more complex in this method, but the improvement in the descriptive power of output maps would be great. As this method would have a simpler structure of initial class definitions, a considerable amount of the complexity and subjectivity would be removed from the ground truth data production process. This approach to class differentiation could potentially better replicate the true class compliment of Landsat pixels as a vast number of ambiguous pixels were excluded from my study's combined ground truth dataset due to their lack of clear membership to any class definition. A second tier classification framework which accounts for all of the permutations of mixed pixels possible (with a two or three level class density modifier) for all land cover types would be able to provide a more accurate estimate of the true size and spatial arrangement of functional land cover classes.

Investigation into Sample Size Requirement

A simple linear regression was carried out to test for the existence of a relationship between ground truth class sample size and classification accuracy. The results of this did not show any clear and significant relationship despite the intuitive nature of the hypothesis. It is likely that there are other factors contributing to the observed variation in each class's classification accuracy.

The lack of support for the hypothesized relationship allows for alternative hypotheses to be formed. These results could be interpreted as indicating that the classification accuracy of the classes defined in the final classification framework each have a different relationship to increasing ground truth sample size. As these classes were derived from a classification framework that defined classes by difference in vegetation species and ecological function, it would be logical to propose that the sources of variation in classification accuracy would be different for each class. There should be no expectation that a classifier would be capable of discriminating these different classes with comparable accuracy from ground truth data pools of equal size. Further investigations into the effects of sample size and quality on classification accuracy of ecologically functional classes need to be carried out to find reliable trends to improve ground truth data production methods and consequently classification accuracy.

Ground Truthing Historic Satellite Data

One of the most important findings of this research was that there is currently no reliable method for producing ground truth data suitable for accuracy assessment of classified maps produced from this historic satellite imagery. This question went unaddressed in the early stages of this research as the original aim was to apply a pre-trained classifier to the historic satellite imagery and a training dataset was not required. There is no intuitive solution to this problem, but it will be important to solve this to provide confidence in the interpretation of classified maps of historic aerial imagery.

One solution would be to draw ground truth data through visual analysis of aerial imagery contemporary with the historic satellite image. This would require the acquisition of archived aerial photography i.e. that which is not old enough to be available through historic aerial photography portals (retrolens.nz or canterburymaps.govt.nz) but too old to be available through the currently accessible council and LINZ data portals. These data have a suite of problems that would need to be overcome before they could be considered reliable enough to be used as a source of ground truth. The greatest problem is the image quality, the older the image the lower the radiometric and spatial resolutions and the harder it will be for a photogrammetry analyst to reliably distinguish the functionally important class details (e.g. remnant/regenerating forest, class density). Furthermore, images are likely to require pre-processing (e.g. orthorectification, mosaicking) and images acquired before c. 1990 are likely to only exist in greyscale. Alternative solutions of assessing classification accuracy of maps produced by application of pre-trained classifiers to historic satellite data may prove more standardisable, especially as the quality of aerial imagery decreases back in time.

Field Surveys

The ground truthing surveys were carried out before investigation into the classification model began due to constraints in the thesis time period and seasonal safety concerns. This meant that the primary aim of the ground truthing surveys was to collect as much data as possible that fit loosely into the classes set out in the target classification framework (Appendix A). Over the course of this research the quality of ground truth sample data was found to strongly influence classification accuracy. This suggests that future research projects will have abundant opportunities to improve on the field survey methods used in this thesis.

Field surveys should be designed to provide a GIS analyst with the data necessary to produce a combined ground truth dataset that will allow for optimal performance of the classifier. Choosing the primary satellite data and secondary ground truth data in advance of the field surveys would enable the surveyor to reproduce the pixel-tracing grid on paper (or mobile device) to assist with boundary drawing in the field. Vegetation should be described as fully as possible to ensure that the primary ground truth dataset contains good reference material should the class definitions change for any reason.

5.3 DEM Production

Method

Aspects of the method described in Shepherd and Dymond (2003) were used to improve the method used in this thesis (section 3.3). The Shepherd and Dymond method included smoothing and resampling steps that mitigated the risk of producing artefacts in the DEM which likely would have translated into errors in the output surface reflectance values. This helped to control for the assumption that the ArcMap “Topo To Raster” tool produced DEMs at a quality suitable for use in the extended chains of calculations used by ATCOR 3 to produce topographically corrected imagery.

DEM Extension

The DEM extension portion of Workflow B (the left half of the flowchart) was necessary because of the following issues identified in ArcMap and ATCOR 3. The ArcMap tool “Topo To Raster” necessarily truncated the output extent of the DEM to a quadrilateral fitting the edges of the contour shapefile despite any attempt to set a larger output extent manually (this was a recurrent problem in scenes with significant areas of sea). Furthermore the “Topo to Raster” tool was very computationally intensive and became somewhat unstable when used to create rasters with too many rows and columns i.e. when either the DEM extent was too large or the DEM output raster cell size was too low. ATCOR 3 would only accept DEM rasters that were larger than the scene boundary. DEMs that did not entirely cover the scene extent were not accepted by ATCOR 3 and the correction process was aborted.

The DEM extension method conformed truncated DEMs to a format accepted by ATCOR 3 through an ad hoc method that could pose risks to the reliability of corrected satellite data if applied without care. The extension method consisted of joining a specifically fabricated raster fragment to the edge of any DEM that did not cover the Landsat scene entirely. Incomplete coverage occurred for two reasons: the secondary scene was slightly offset due difference in acquisition date or the ArcMap “Topo to Raster” algorithm was not able to fill pixels with data where there were areas of sea regardless of the status of processing extent settings before execution. This meant that DEMs had to be extended for all three secondary satellite datasets.

Correction issues may arise if DEM extension is carried out in a region of the map too close to the AOI. The atmospheric and topographic correction algorithms used by ATCOR 3 are complex and were not comprehensively examined in this research, but the portion of a pixel’s reflectance that is calculated to originate from diffuse illumination (i.e. light that reaches the subject after reflection from sloped land surfaces and their land covers) may be erroneously affected by the method of DEM extension described in the methods section if it is close enough to the AOI to potentially influence a pixel's calculated "skyview" value (Dymond et al., 2001, Richter & Schläpfer, 2019). As all pixels in the extension raster segment were set to a value of 1, there is a chance that this could have altered the results of ATCOR 3's correction algorithms. This may have been a problem if the AOIs had been near the edge extension data, but in all cases, the extensions were hundreds of kilometres away and any changes to the topographically corrected spectral values would have been relatively localised.

A better method of achieving this for larger scale use would be to set up custom sensor parameters in ATCOR 3 that allow for input of custom sized imagery and DEMs clipped to match the AOI extent. This would have mitigated the risk of data corruption from fabricated raster fragments and would have allowed for much faster processing during both the corrections phase (which would have been a minor advantage) and the DEM production phase (which would have been a great advantage). This would have required a much deeper knowledge of the MODTRAN parameters and the localised environmental parameters to achieve and this more thorough method was not able to be implemented within the time constraints of this thesis.

5.5 Satellite Image Correction

Correction of satellite imagery from "at sensor" digital number values to "ground level" surface reflectance values was partially successful. The images appeared obviously flattened on visual examination (indicating success in correcting for the topographic effect) but some artefacts were produced and some problems in uniformity between primary and secondary satellite datasets were identified. The correction process described in Workflow C1 did not create ideal data products and is likely to be a suboptimal optimal method for correcting Landsat imagery for time series analysis of land cover.

Two of the primary satellite images suffer from artefacts created in the correction process. These appear as white, cloudlike objects over dark areas of forest in corrected Landsat 8 images. These artefacts do not appear in the corrected images of the Canterbury AOI, perhaps due to the significantly smaller area of dark forest and lower overall abundance of steep land. These artefacts are, however, common and cover large areas in images of the Manawatū-Whanganui AOI and the Northland AOI. They likely occur due to an overcompensation in the topographic correction process as they are only visible on south and west facing slopes where shadows would be present and are not visible on examination of the source data.

Specific artefacts were also noticed in all secondary satellite datasets in the form of banding over areas of sea. This appears to be caused by a digital imaging error as it is visible in the source data, the bands are visible at regular intervals and are they are aligned with the image frame. This variation is difficult to see over land, but as it is uniformly observed across all datasets in regions where the surface reflectance is regular, these artefacts are assumed to affect the spectral values of all pixels in the image. Additionally, all three secondary satellite images appeared noisy at high magnification and characteristic edges of land cover are often much less distinct. This is also likely to be signal independent noise and may originate from inaccuracy in the Landsat 4 digital sensor or the method of quantizing the reflectance data to 8-bit pixels.

Future image correction should focus on appropriate strength of topographic correction to ensure artefacts are not introduced to the data during this process. The need for filtering of the secondary satellite datasets is also of great importance and is demonstrated by the observed banding across all three secondary satellite images. Introduction of both these sources of variation into the datasets will have caused inaccuracies to develop in the classifier model and in the secondary satellite datasets which prevented accurate detection of land cover classes on classifier application.

Extensive trial and error work with different corrections techniques proved inconclusive in increasing the back translatability of trained classifier models. This problem would be improved by the development of a method of attaining accurate ground truth data for the secondary satellite images (similar to the problems in ground truth data production described on p. 96). Statistical differences in the spectral values of class samples taken from primary and secondary satellite datasets could be compared to measure the degree of agreement produced by a particular correction process. Any measured differences could be used to produce a key transformation that improves comparability of the two datasets. The most modern Landsat 4/5 TM data and some of the early Landsat 7 ETM+ data are likely to have contemporary aerial imagery available as a source of ground truth data, although the suitability of this data for this purpose is unknown.

The most obvious difference between the original primary and secondary satellite datasets is the radiometric resolution. The Landsat 8 OLI/TIRS sensor records data at 12 bits per pixel while the Landsat 4 TM sensor recorded data at 8 bits per pixel. Applying class definitions created from data with a high dynamic range to a dataset with lower dynamic range is likely to be a be a source of some error in the classified maps of secondary satellite data. Both these datasets were normalised during the image correction process to 16-bit unsigned integer data, this transformation was assumed to be equitable and output data was assumed to be comparable. This transformation was an opaque internal operation of the ATCOR 3 software however and the effect of this manipulation on class comparability between primary and secondary satellite datasets was not tested.

Software selection for image correction was critically important, and the final decision of which software to use for satellite image correction was not made until the last quarter of the work period as the investigation of the most effective approach proved to be a great challenge. ATCOR was chosen due to its wide variety of useful tools and its relatively user friendly and stable software package. It is rigid in its input parameters however, and to get the most out of this software would require a considerable amount of research and preparation. Alternatively, a custom method could be developed that more closely resembles the WAK II model described by Shepherd and Dymond (2003). This method is built on top of the 6S radiative transfer model, an alternative to the MODTRAN radiative transfer model which underlies ATCOR. There is some evidence to suggest that models build on the 6S code may outperform MODTRAN based models (Nazeer et al., 2014) when classifying land covers from Landsat data. More research is required to determine the optimal image correction approach for the specific task of classifying land cover from Landsat imagery considering New Zealand's atmospheric and topographic environment.

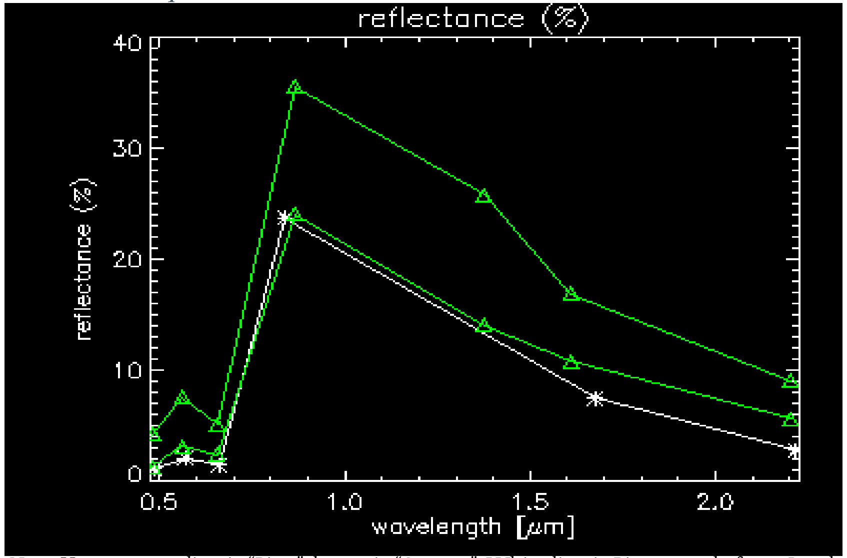
"SPECTRA" module

The "SPECTRA" module in ATCOR 3 allowed comparison of estimated surface reflectance values with a library of spectral profiles of examples of different land cover classes. This is a useful tool for testing the accuracy of a specific image correction process. Although the spectral library packaged with the ATCOR software does not contain samples that conform to the classification frameworks used in this research, this coarse assessment of correction success has potential to be a useful diagnostic tool.

One option for further research into this area would be to build a custom spectral library applicable to the specific land cover classes in the classification framework. This would allow for easy comparison between different MODTRAN atmospheric file templates to find the most accurate method of attaining true surface reflectance values of the land cover classes of interest. If this tool showed that none of the corrections performed adequately then there would be good evidence for developing a custom corrections model through similar methods to Shepherd and Dymond (2003).

Although the spectral profiles in the ATCOR reference library were not ideal matches for New Zealand Land cover classes some examples of spectral profile comparisons are shown in figures 38-41. Some trends can be seen that informed the decisions made in determining the most appropriate corrections parameters for this research. Figures 38-41 are clipped directly from the ATCOR 3 "SPECTRA" module. Green lines are the spectral library samples provided by ATCOR 3, white and yellow lines represent the average spectral profile of a 5x5 pixel sample of a selected patch of continuous land cover. Samples were selected from a range of points, the samples chosen for display are representative of a typical spectral profile from a selection of 10 chosen samples of a land cover class. Although the values of the library samples may differ in magnitude from the actual values of chosen image samples, the trend of the spectral profile is more important for providing confidence in the accuracy of the correction (Richter & Schläpfer., 2019). Only the best matching library spectra were chosen which is why the “Pine” and “Spruce” samples are present in most of the charts.

Figure 38. Spectral profile of a 2014 Canterbury Pine sample compared with "Pine" and "Spruce"



Note. Upper green line is "Pine", lower is "Spruce". White line is Pine sample from Landsat data.

Figure 39. Spectral Profile of a 2014 Canterbury Pasture Sample Compared With "Meadow"

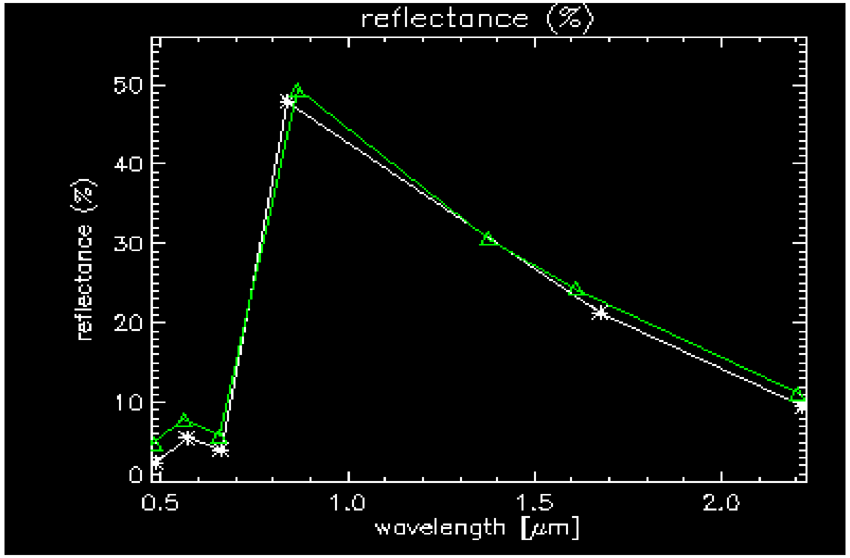
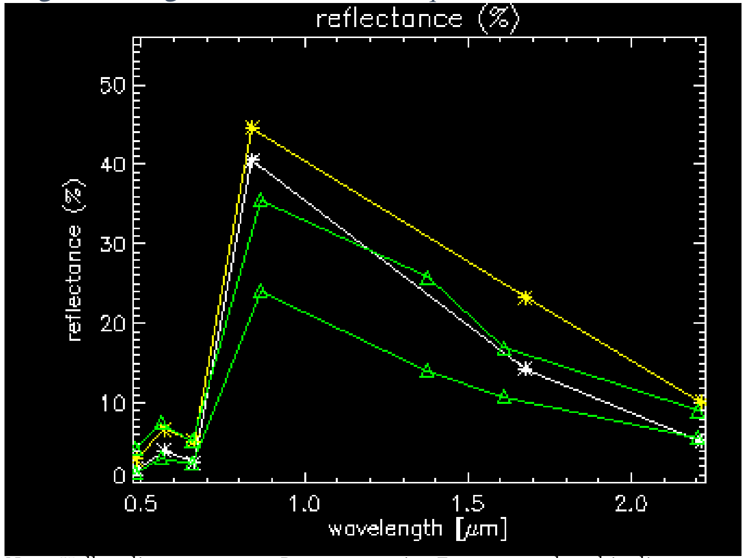
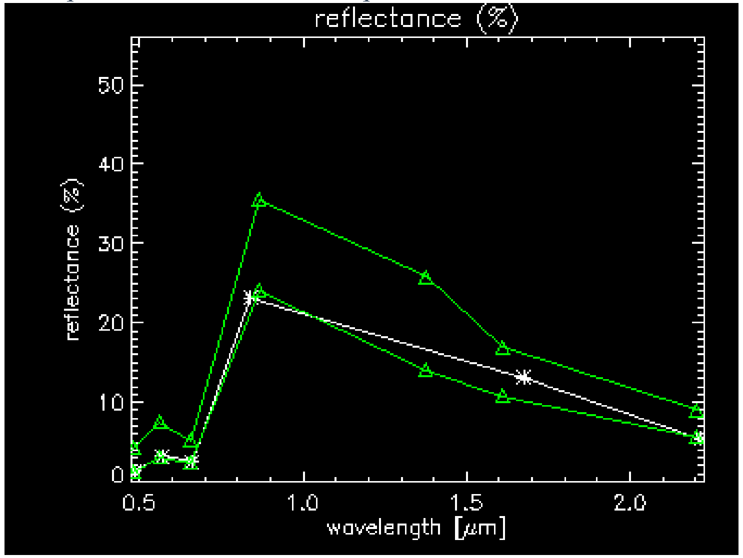


Figure 40. Spectral Profile of a 2014 Canterbury Remnant Native Forest and Regenerating Native Forest Compared With "Pine" and "Spruce"



Note. Yellow line represents Remnant native Forest sample, white line represents Regenerating Native Forest sample

Figure 41. Spectral Profile of a 2014 Canterbury Remnant Kanuka Forest Compared With Pine and Spruce



Note. Upper green line is "Pine", Lower is "Spruce". White line is continuous kanuka forest sample from Landsat data.

Topographic Correction

The Topographic correction methods applied to Landsat data in this research were chosen due to evidence in the literature of the greater effect of Modified Minnaert algorithms on NIR and SWIR1 and SWIR 2 bands (Richter et al., 2009; Park et al., 2017, Young et al., 2017). These bands are where variations in the spectral profile of vegetation is most pronounced and therefore where the best discriminability of vegetative land cover classes is likely to be found.

Investigation and testing of the manually adjustable parameters of topographic correction in ATCOR 3 was not comprehensive. In general, pre-set parameters and those calculated by ATCOR during image processing were used. Limited experiments were carried out to test the effect of changing some topographic correction parameters on the training and application of classifiers, but due to the length of the workflow required to test each change independently, it was impossible to thoroughly test all the options available. There is an opportunity here for future work to focus on atmospheric and topographic correction methods and how best to customise these to deal with the unique challenges that are posed in automated classification of Landsat data in New Zealand.

Proposed Pre-Processing for Future Research

An appropriate model for interpolation and smoothing of secondary satellite data or resampling of primary satellite data to allow for better comparability of values was not developed. The images used in this study from 1989 and 1990 suffer from visually obvious data noise and artefacts that are signal independent and not present in the primary satellite images. A data smoothing algorithm may be advantageous in allowing a classifier trained on clearer imagery to better recognise this data. Basic gaussian filters were applied through eCognition but were unsuccessful in improving the classifier's ability to recognise the chosen land cover classes in older imagery.

5.6 Classification

Masking

Water and bare ground were able to be classified through an index threshold method with good visual accuracy. Improvements in classification accuracy of all other classes were immediately apparent after water and bare ground classes were masked out and removed from the classifier model. This trend of better classification accuracy with fewer model classes was observed in this research throughout the classifier and classification framework stages.

Hybrid classifier model

Phiri and Morgenroth 2017 suggest that hybrid methods have promise for land cover classification of Landsat imagery and that this requires more research. The classifier used in my research was a hybrid method containing two separate classifier models. One CNN classifier formed the basis of class separation by producing probability heatmap rasters for each class, and one “Random Trees” classifier used ground truth samples to classify image objects based on the probability values of each class’s heatmap layer. This classifier model was effective at predicting some land cover classes and ineffective at predicting others. Although the output classified maps were not highly accurate, the overall classifier process structure is stable and easy to manipulate which will allow it to be widely applicable to a range of multispectral image classification tasks.

The CNN used in my research is a modified version of Google’s open source TensorFlow platform (Trimble., 2019). It is constrained and implemented through a process tree algorithm within eCognition and is operated through the programme’s friendly graphical user interface. Neural network algorithms are often described as “black boxes” due to their inherently unexaminable internal processes. The features used to define each class are parameterised in “hidden layers” of the CNN, so there are relatively very few parameters available to the user to adjust. Of the parameters that are available, the process of finding the most optimal solution with the highest classification accuracy was carried out through iterative testing of each parameter in isolation. An exhaustive development cycle may also examine the covariant, correlational and mixed effects of adjusting each parameter to attain an optimal model, but the development of a stable and flexible classifier platform was the focus of this research and exhaustive optimisation was not carried out.

The most important limitation in this model is its reliance on ground truth data. Every erroneously classified pixel in the combined ground truth dataset contributed to uncertainty in the model and reduced clarity in the suite of spectral and spatial features used to determine the CNN's definition of each class. This classifier requires accurately classified input data to produce accurately classified output maps.

The edges of all classified maps should be removed from future analyses due to classification artefacts produced by the CNN's convolution kernel. These did not occur in early machine learning trials but were pronounced in early CNN trials where the classifier produced a suboptimal result. Future studies should ensure that AOIs have an appropriate buffer to allow edge artefacts on classified maps to be removed before analysis.

Application of Trained Classifier to Historic Satellite Imagery

The classification method developed in my research was not able to produce classified maps with any accuracy useful for measuring land cover status prior to the introduction of the LCDB in 1996. Much of the classified area of these maps bore no resemblance to the satellite image but some land cover boundaries were defined with reasonable visual accuracy. For example, areas of known continuous native remnant ('a1c') in the secondary image of the Manawatū-Whanganui AOI were visible although they were almost entirely classified incorrectly as continuous planted exotic ('b1c'). A similar pattern of substitution can be seen (although less clearly) in the classified map of the secondary image of the Northland AOI. Although this does not allow any useful estimations of historic NWV distribution, the detection of patch edges is a promising sign for future classifier development. Although this classifier model is likely to still require some optimisation to exclude any chance of classifier error in the final map products, the improvements observed during development indicate that input data (ground truth and satellite imagery) quality is the limiting factor in achieving high classification accuracy.

Accuracy Assessment

The accuracy assessment tool in eCognition calculates five of the most commonly applied accuracy assessment indices for interpretation of classified map outputs. My research focused on the differences in user and producer accuracies and used the Kappa Index of Agreement (KIA) to quantify the significance of accuracy measurements by subtracting the probability of accurate classification occurring by chance from the accuracy rate. The difference in a class's user and producer accuracy provides a remote sensing analyst with a better picture of the trend of misclassification in a class as well as the approximate magnitude of accuracy. Where a class has much higher user accuracy than producer (and the KIA is high), there is a partial indication that the area of this class shown in the map is larger than the actual area (the class may be over classified). Where the reverse is true then there is a partial indication that the area of this class shown on the map is smaller than the actual area (the class may be under classified). While this metric is coarse and cannot provide any significant conclusions on its own, it can direct the remote sensing analyst's efforts in including or excluding ambiguous pixels from a class in the ground truth dataset.

The ground truth data selection process used in this thesis was vulnerable to bias due to variability in the selector's judgement. Approximately 2000-4000 pixels were used in each satellite image's combined ground truth dataset and the contents of each was checked by a human observer at least once. The likelihood of selection errors occurring in a human filtered dataset this large are high and the accuracy assessment tool was used to direct efforts in ground truth data selection towards those classes that performed especially poorly or had a large spread between user and producer accuracy. The effect of this subjectivity problem on classification accuracy of low density classes may be a reason to investigate development of a continuous only classification scheme so that a more complete and complex array of low density land cover classes may be calculated from the probability heatmaps of well-defined continuous classes.

6. Conclusions

6.1 Key Outcomes

Although some of the original objectives of this research were not achieved, the information presented in this thesis should allow future researchers to more easily define and execute methods for estimating attributes of past land covers in agroecosystems. The key outcomes of this research are presented below as a summary of the objectives achieved and the future objectives identified.

Objectives Achieved

1. A simple method for creating detailed ground truth datasets for classification of modern satellite imagery by combining observations from aerial imagery and ground surveys (the pixel-tracing method section 3.2).
2. An effective machine learning classifier built and trained to identify a framework of New Zealand's woody vegetation landcover classes with functional relevance to important environmental features.
3. A classification framework for woody vegetation on sheep and beef farmland was developed that accounts for the current limits of discriminability of different vegetation classes and densities in Landsat 8 OLI/TIRS data while including a high level of environmentally relevant detail.

Future Objectives Identified

1. Development of a classification framework that is based on statistical separability of spectral and spatial attributes of landcover classes to optimise classification accuracy. This must be conformable to the functional classification of woody vegetation in agroecosystems suggested by Forbes et al. (2020) for presentation.
2. Development of a method for estimating classification accuracy of maps produced by application of a trained classifier to satellite data with no (or very poor quality) ground truth available.
3. More development in ground truth data production methods. Development of targeted survey methods to allow optimal classification accuracy of poorly classified classes.
4. Research into methods of noise reduction and bit depth standardisation for Landsat data produced by sensors older than Landsat 8.

There is a disconnect between the two concepts: all the classes discriminable in the data and all the functional classes of New Zealand woody vegetation. This needs to be bridged in order to progress any of the other questions that could be applied to the future objectives. The methods by which these two classification frameworks are aligned will cause specific bias in the results so they must be chosen carefully.

Ground truth data quality was the most important variable in improving classification accuracy in the classified maps of the primary satellite datasets. The specific parameters that define quality in this research are detailed in the methods (section 3.2, p. 22), but these must be underpinned by a classification framework of detailed class definitions that minimise the chances of the analyst encountering ambiguous pixels.

Continuous cover classes were consistently observed to have higher classification accuracy than diffuse cover classes. The spectral variability and frequency of ambiguous pixels in the ground truth datasets of diffuse classes are likely to be much higher than continuous classes. A thorough statistical analysis of the spectra of poorly classified classes must be underpinned by a clearly defined classification framework.

Data normalisation and noise reduction are a second key point that will be necessary for successful backwards application of a trained classifier. Atmospheric and topographic corrections must be carried out with care and additional normalisation methods devised to deal with the variability caused by differences in pixel bit depth and data noise in historic imagery.

Areas of continuous remnant native vegetation were classified with relatively good accuracy overall considering their visual similarity to continuous regenerating native and continuous planted exotic classes in the satellite data. Further research into spectral separability of this class and development of a classification framework that accounts for the results may allow highly accurate discrimination of remnant native forest from Landsat data.

Accuracy assessment of land cover maps of historic data created by application of machine learning classifiers trained with modern ground truth data is impossible through traditional methods. This problem must be solved to support any claims of reliability made in future, more successful classification of historic satellite imagery.

While this study was unable to achieve many of its goals, the barriers to reaching them are now clearer. This thesis should be used as a guide for developing methods of satellite data processing and classification suitable for analysing Landsat imagery acquired before the introduction of LCDB in 1996 so that they are directly comparable with modern classified Landsat data.

7. References

Beef and Lamb New Zealand. (2018). Environment Strategy and Implementation Plan 2018-22. Retrieved March 2019 from: https://beeflambnz.com/sites/default/files/news-docs/Env-strategy-Imp-plan.pdf on 10/6/18

Beef and Lamb New Zealand. (2018). Farm Classes. Retrieved March 2019 from: https://beeflambnz.com/data-tools/farm-classes

Chang, T., & Kak, A. C. (2019). Mining concise datasets for testing satellite-data-Based land-cover classifiers meant for large geographic areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(9), 3562-3577. https://doi.org/10.1109/jstars.2019.2935185

Cieraad E, Walker S, Price R, Barringer J 2015. An updated assessment of indigenous cover remaining and legal protection in New Zealand's land environments. New Zealand Journal of Ecology 39, 309-315.

Congalton, R. G., & Green, K. (2002). *Assessing the accuracy of remotely sensed data: Principles and practices.* Boca Raton, *FL: CRC Press.*

Deconchat, M., Brockerhoff, E., & Barbaro, L. (2009). Effects of surrounding landscape composition on the conservation value of native and exotic habitats for native forest birds. *Forest Ecology and Management*, *258*, S196-S204. doi:10.1016/j.foreco.2009.08.003

Dymond, J., Shepherd, J., & Qi, J. (2001). A simple physical model of vegetation reflectance for standardising optical satellite imagery. *Remote Sensing of Environment*, *77*(2), 230-239. https://doi.org/10.1016/s0034-4257(01)00288-7

ESRI (2020). *Indices gallery*. Retrieved 9th March 2020 from: https://pro.arcgis.com/en/pro-app/help/data/imagery/indices-gallery.htm

Ewers, R. M., Kliskey, A. D., Walker, S., Rutledge, D., Harding, J. S., & Didham, R. K. (2006). Past and future trajectories of forest loss in New Zealand. *Biological Conservation*, *133*(3), 312-325. doi:10.1016/j.biocon.2006.06.018

Forbes, A., Norton, D., Buckley, H., Case, B., & Pannell, J. (2020). A functional classification of woody vegetation in agroecosystems. Unpublished.

Franklin, J. F., & Lindenmayer, D. B. (2009). Importance of matrix habitats in maintaining biological diversity. *Proceedings of the National Academy of Sciences*, *106*(2), 349-350. doi:10.1073/pnas.0812016105

Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. Journal of Plant Physiology, 161(2), 165-173.

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. Remote Sensing of Environment, 80(1), 76-87. https://doi.org/10.1016/S0034-4257(01)00289-9

Gouin, D. (2006). *Agricultural sector adjustment following removal of government subsidies in New Zealand* (284). Retrieved March 2019 from: https://researcharchive.lincoln.ac.nz/bitstream/handle/10182/171/aeru_rr_284.pdf?sequence=1&isAllowed=y Ministry for the Environment. (2009). Land Environments New Zealand (LENZ) – Level 2. Retreived from: https://data.mfe.govt.nz/layer/52357-land-environments-new-zealand-lenz-level-2-polygons-2009/ on 10/6/18

Huete, A. (1988). A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment, 25(3), 295-309. https://doi.org/10.1016/0034-4257(88)90106-x

IDB - Index DataBase. (n.d.). *Landsat 8 indices*. Retrieved 9th March 2020 from: https://www.indexdatabase.de/db/is.php?sensor_id=168

Landcare Research New Zealand Ltd. (2020, January 29). *LCDB v5.0 - Land cover database version 5.0, Mainland New Zealand*. LRIS Portal. Retrieved 28th April 2020 from: https://lris.scinfo.org.nz/layer/104400-lcdb-v50-land-cover-database-version-50-mainland-new-zealand/metadata/

Land Information New Zealand (LINZ). (2019). *NZ Aerial Imagery*. Canterbury 0.3m Rural Aerial Photos. (2014-2015). Sourced from the LINZ Data Service and licensed by The Canterbury Aerial Imagery (CAI) consortium for re-use under the Creative Commons Attribution 4.0 International. Retrieved April 2019 from: https://data.linz.govt.nz/layer/52602-canterbury-03m-rural-aerial-photos-2014-2015/

Land Information New Zealand (LINZ). (2019). *NZ Aerial Imagery.* Manawatū-Whanganui 0.3m Rural Aerial Photos. (2016-2017). Sourced from the LINZ Data Service and licensed by **(**Aerial Surveys Ltd and Manawatū-Wanganui LASS Limited**)** for re-use under the Creative Commons Attribution 4.0 International. Retrieved April 2019 from: https://data.linz.govt.nz/layer/88145-Manawatū-whanganui-03m-rural-aerial-photos-2016-2017

Land Information New Zealand (LINZ). (2019). *NZ Aerial Imagery.* Northland 0.4m Rural Aerial Photos. (2014-2016). Sourced from the LINZ Data Service and licensed by Northland Aerial Imagery Consortium (NAIC) for re-use under the Creative

Commons Attribution 4.0 International. Retrieved April 2019 from: https://data.linz.govt.nz/layer/88131-northland-04m-rural-aerial-photos-2014-2016/

Land Information New Zealand (LINZ). (2019). NZ Contours (Topo, 1:50k). (2011). (File Identifier - 4e287c2e-fb6c-ea67-6179-36facda52606). (2019). Retrieved August 12, 2019 from Land Information New Zealand website: https://data.linz.govt.nz/layer/50768-nz-contours-topo-150k/

McGlone M. S. (1983). Polynesian deforestation of New Zealand: A preliminary synthesis. Archaeology in Oceania, 18(1). 11-25.

Ministry for the Environment. (2017). A Guide to the National Policy Statement for Freshwater Management 2014 (as amended 2017). Wellington: *Ministry for the Environment*. Retrieved November 20th 2019 from: https://www.mfe.govt.nz/sites/default/files/media/Fresh%20water/nps-fm-guide-2017-final.pdf

Ministry for the Environment. (2019). Action on agricultural emissions: A discussion document on proposals to address greenhouse gas emissions from agriculture. Wellington: *Ministry for the Environment*. Retrieved November 20th 2019 from: https://www.mfe.govt.nz/sites/default/files/media/Climate%20Change/action-on-agricultural-emissions-discussion-document.pdf

Ministry for the Environment. (2019). He Kura Koiora i hokia: A discussion document on a proposed National Policy Statement for Indigenous Biodiversity. Wellington: *Ministry for the Environment*. Retrieved December 2nd 2019 from: https://www.mfe.govt.nz/sites/default/files/media/Biodiversity/he-kura-koiora-i-hokia-discussion-document.pdf

Ministry for the Environment. (2019). *Land Environments New Zealand (LENZ) – Level 1 Polygons*. (2009). Retrieved February 15th 2020 from: https://data.mfe.govt.nz/layer/52354/

Ministry for the Environment. (2018). Our Land 2018. New Zealand's Environmental Reporting series. Publication reference number: ME1350. Retrieved 10th June 2018 from: http://www.mfe.govt.nz/sites/default/files/media/RMA/Our-land-201-final.pdf

Monks, A., Hayman, E., & Walker, S. (2019). Attrition of recommended areas for protection: clearance of ecologically significant vegetation on private land. *New Zealand Journal of Ecology*, *43*(2). https://doi.org/10.20417/nzjecol.43.15

Nazeer, M., Nichol, J. E., & Yung, Y. (2014). Evaluation of atmospheric correction models and Landsat surface reflectance product in an urban coastal environment. *International Journal of Remote Sensing*, *35*(16), 6271-6291. https://doi.org/10.1080/01431161.2014.951742

Norton, D., Suryaningrum, F., Buckley, H., Case, B., Cochrane, H., Forbes, A., & Harcombe, M. (2020). Achieving win-win outcomes for pastoral farming and biodiversity conservation in New Zealand. *New Zealand Journal of Ecology*, *44*(2). https://doi.org/10.20417/nzjecol.44.15

Norton, D. and Pannell, J. (2018). Desk-top assessment of native vegetation on New Zealand sheep and beef farms. *Beef and Lamb New Zealand Ltd*.

Park, S., Jung, H., Choi, J., & Jeon, S. (2017). A quantitative method to evaluate the performance of topographic correction models used to improve land cover identification. *Advances in Space Research*, *60*(7), 1488-1503. https://doi.org/10.1016/j.asr.2017.06.054

Perry, G. L., Wilmshurst, J. M., McGlone, M. S., McWethy, D. B., & Whitlock, C. (2012). Explaining fire-driven landscape transformation during the Initial Burning Period of New Zealand's prehistory. *Global Change Biology*, *18*(5), 1609-1621. doi:10.1111/j.1365-2486.2011.02631.x

Phiri, D., & Morgenroth, J. (2017). Developments in Landsat land cover classification methods: A review. *Remote Sensing*, *9*(9), 967. doi:10.3390/rs9090967

Phiri, D., Morgenroth, J., & Xu, C. (2019). Four decades of land cover and forest connectivity study in Zambia—An object-based image analysis approach. *International Journal of Applied Earth Observation and Geoinformation*, *79*, 97-109. doi:10.1016/j.jag.2019.03.001

Raschka, S. (2020). *What is the difference between a parametric learning algorithm and a nonparametric learning algorithm?* Dr Sebastian Raschka. Retrieved 3rd February 2020 from: https://sebastianraschka.com/faq/docs/parametric_vs_nonparametric.html

Richardson, A. J., & Weigand, C. (1977). Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing*, *43*(12), 43.

Richter, R., Kellenberger, T., & Kaufmann, H. (2009). Comparison of topographic correction methods. *Remote Sensing*, *1*(3), 184-196. https://doi.org/10.3390/rs1030184

Richter, R., and Schläpfer, D. (2019). *ATCOR-2/3 User Guide, Version 9.3.0, October 2019.* Wessling, Germany: DLR – German Aerospace Center, Wil SG Switzerland: ReSe

Applications LLC. Retrieved December 20th 2019 from: www.rese-apps.com/software/download

Rocchio and Barsi. (2017) retrieved December 16th 2019 from: https://www.usgs.gov/media/images/bandpass-wavelengths-all-landsat-sensors

Sharma, A., Liu, X., Yang, X., & Shi, D. (2017). A patch-based convolutional neural network for remote sensing image classification. *Neural Networks*, *95*, 19-28. https://doi.org/10.1016/j.neunet.2017.07.017

Shepherd, J. D., & Dymond, J. R. (2003). Correcting satellite imagery for the variance of reflectance and illumination with topography. *International Journal of Remote Sensing*, *24*(17), 3503-3514. https://doi.org/10.1080/01431160210154029

Song, H., Kim, Y., & Kim, Y. (2019). A patch-based light Convolutional neural network for land-cover mapping using landsat-8 images. *Remote Sensing*, *11*(2), 114. https://doi.org/10.3390/rs11020114

Spectral Sciences Inc. (n.d.). *MODTRAN®*. MODTRAN®. Retrieved March 2019 from: https://modtran.spectral.com/modtran_faq

Taylor, R., Smith, I., Cochrane, P., Stephenson, B., & Gibbs, N. (1997). The State of New Zealand's Environment*. The Ministry for the Environment*. Retreived March 2019 from: https://www.mfe.govt.nz/sites/default/files/ser-1997.pdf

Trimble. (2019) eCognition Developer 9.5 User Guide Version 9.5.1. *Trimble Germany GmbH.*

USGS – U.S. Geological Survey. (2020). *A Landsat timeline « Landsat science*. Landsat Science. Retrieved May 2nd 2020 from: https://landsat.gsfc.nasa.gov/a-landsat-timeline/

USGS - U.S. Geological Survey. (n.d.). EarthExplorer - Home. Retrieved March 1st, 2019, from: https://earthexplorer.usgs.gov/

USGS - U.S. Geological Survey. (n.d.). Landsat 7. Retrieved January 17th 2020 from: https://www.usgs.gov/land-resources/nli/landsat/landsat-7?qt-science_support_page_related_con=0#qt-science_support_page_related_con

USGS – U.S. Geological Survey. (2019). Landsat Collection 1 level 1 product definition (LSDS-1656 V.2.0). Retrieved January 23rd, 2020 from: https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1656_%20Landsat_Collection1_L1_Product_Definition-v2.pdf

USGS – U.S. Geological Survey. (1997). Landsat data. A brief history of the Landsat program (Fact Sheet 084-97). Retrieved January 23rd, 2020 from: https://pubs.usgs.gov/fs/1997/0084/report.pdf

USGS - U.S. Geological Survey. (n.d.). Landsat Processing Systems. Retrieved January 15th, 2020 from: https://www.usgs.gov/land-resources/nli/landsat/landsat-processing-systems

USGS - U.S. Geological Survey. (2018). *Landsat Surface Reflectance*. USGS.gov | Science for a changing world. Retrieved March 16th 2020 from: https://www.usgs.gov/land-resources/nli/landsat/landsat-surface-reflectance?qt-science_support_page_related_con=0#qt-science_support_page_related_con

USGS - U.S. Geological Survey. (n.d.). National Land Archive Production System. Retrieved January 15th, 2020 from: https://www.usgs.gov/land-resources/nli/landsat/national-land-archive-production-system

USGS - U.S. Geological Survey. (n.d.). USGS EROS Archive - Landsat Archives - Landsat 4-5 TM Level-2 Data Products - Surface Reflectance. Retrieved January 15th, 2020 from: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-landsat-archives-landsat-4-5-tm-level-2-data-products-surface?qt-science_center_objects=0#qt-science_center_objects

U.S. Geological Survey. (n.d.). *What Are the Band Designations for the Landsat Satellites?* USGS.gov | Science for a Changing World. Retrieved January 15th 2020 from: https://www.usgs.gov/faqs/what-are-band-designations-landsat-satellites?qt-news_science_products=0#qt-news_science_products

USGS - U.S. Geological Survey. (2018). What does the USGS do? Retrieved January 15th 2020 from: https://www.usgs.gov/faqs/what-does-usgs-do?qt-news_science_products=0#qt-news_science_products

Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, *2017*, 1-17. https://doi.org/10.1155/2017/1353691

Vittek, M., Brink, A., Donnay, F., Simonetti, D., & Desclée, B. (2014). Land cover change monitoring using Landsat MSS/TM satellite image data over west Africa between 1975 and 1990. *Remote Sensing*, *6*(1), 658-676. https://doi.org/10.3390/rs6010658

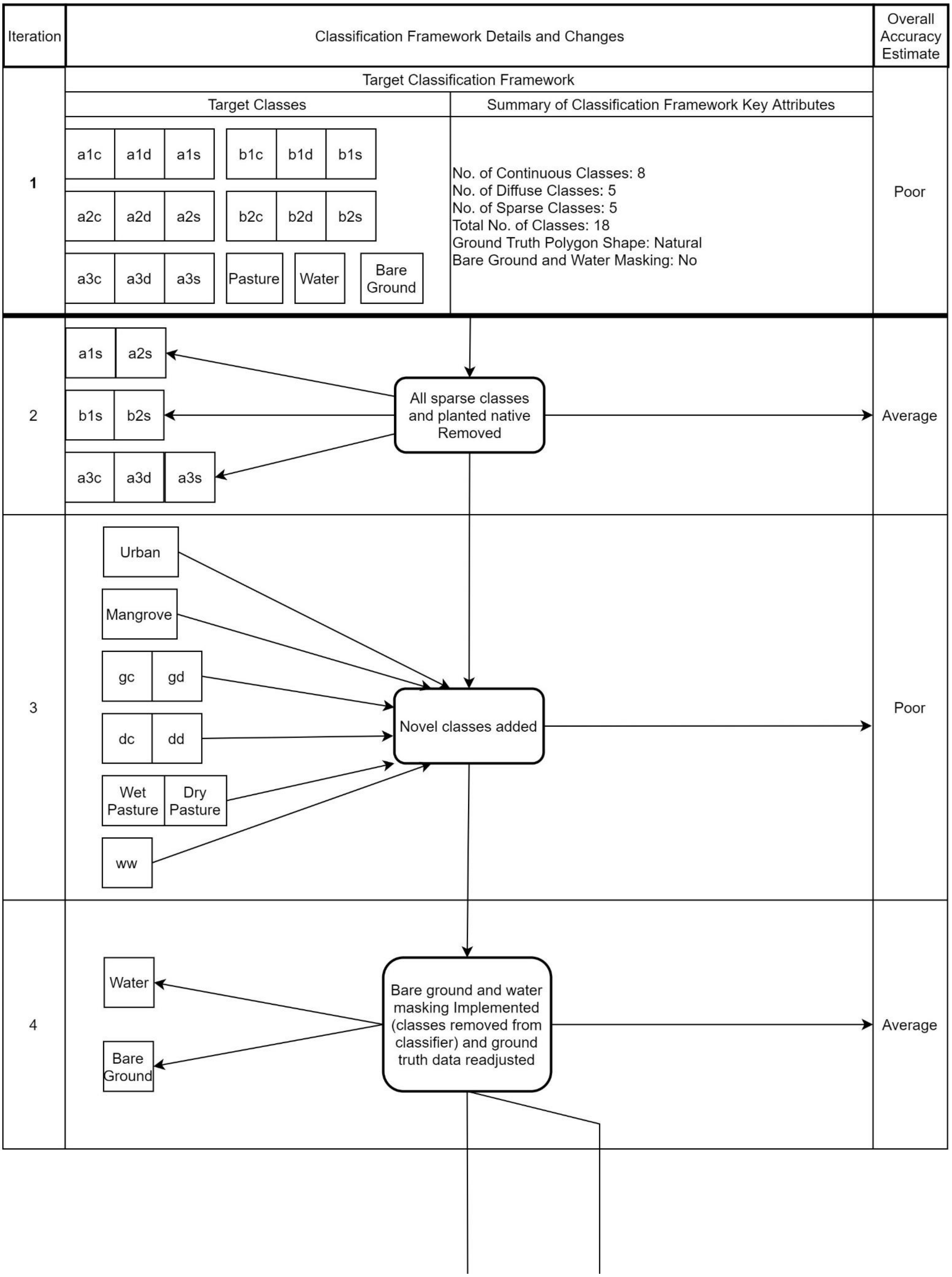
Young, N. E., Anderson, R. S., Chignell, S. M., Vorster, A. G., Lawrence, R., & Evangelista, P. H. (2017). A survival guide to Landsat preprocessing. *Ecology*, *98*(4), 920-932. https://doi.org/10.1002/ecy.1730

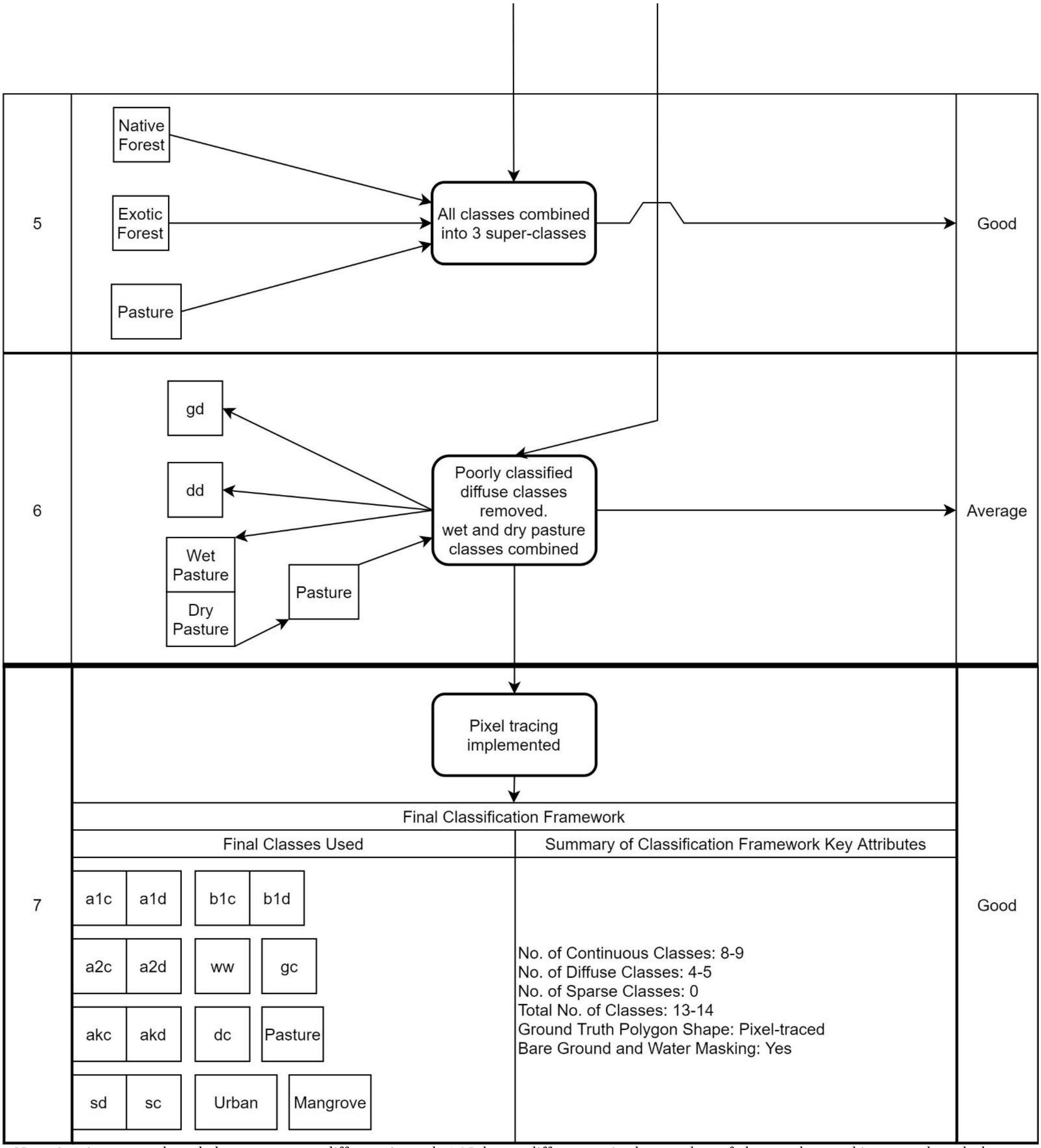
Appendix A: Target Land Cover Classification Framework for Assessment of Change in Woody Vegetation Cover on Sheep and Beef Farms

Code	Class			Definition	Characteristic Taxa
a1c	Native	Remnant	Continuous	Patches present since before human arrival containing old-growth canopy trees. May have been modified by some logging or animal damage.	Elaeocarpus spp., Kauri, Nothofagaceae spp., Podocarpaceae spp., Puriri, Rewarewa, Tawa, Tītoki
a1d	“	“	Diffuse	A degraded form of “alc” class containing old-growth canopy trees as well as grassy clearings	“
a1s	“	“	Sparse	Old growth trees scattered throughout pasture	“
a2c	“	Regenerating	Continuous	Patches of native woody vegetation that have established on land that was previously dominated by exotic vegetation	Coprosma spp., Kahikatea, Kanuka, Mahoe, Manuka, Pittosporum spp., Pseudopanax spp., Totara
a2d	“	“	Diffuse	Similar to “a2c” but with a highly interrupted canopy and an abundance of grassy clearings	“
a2s	“	“	Sparse	Relatively young native woody vegetation growing at low density throughout pasture	“
a3c	“	Planted	Continuous	Patches of planted native woody vegetation such as restoration sites	Any native woody species
a3d	“	“	Diffuse	A possible but unlikely spatial conformation of planted native trees	“
a3s	“	“	Sparse	Scattered native trees. Generally park or garden plantings	“
b1c	Exotic	Planted	Continuous	Plantation forests for timber production or dense sections of shelterbelts/windbreaks	Douglas Fir, Eucalyptus spp., Pinus spp., Populus spp., Salix spp.
b1d	“	“	Diffuse	Typically timber crops planted diffusely amongst pasture used for grazing (agroforestry). Will likely also be used for young plantation forests and some sections of windbreaks/shelterbelts	“
b1s	“	“	Sparse	Sparse exotic trees, often planted for erosion control	“
b2c	“	Regenerating/Weedy	Continuous	Dense patches of weedy self-sown exotic trees	Same as the b1 classes but also including Gorse, Scotch Broom and other non-commercial species
b2d	“	“	Diffuse	Similar to “b2c” but interspersed with grassy clearings	“
b2s	“	“	Sparse	Individuals or small groups of trees amongst pasture or continuous native land covers	“
Pasture	Pasture/Crop	-	Continuous	Agricultural land covered uniformly with pasture or crops	Grasses, Brassicaceae etc.
Water	Water	-	Continuous	Regions where water covers >70% of the area	None
BG	Bare Ground	-	Continuous	Regions where >70% of the area is bare of vegetation	None

Note. Density class key: Continuous= >70% coverage; Diffuse= 15-70% coverage; Sparse= <15% coverage. Modified from Forbes et al. (2020).

Appendix B: Summary of Classification Framework Development





Note. Continuous and total class counts are different in each AOI due to differences in the number of classes observed in ground truth data. Canterbury did not contain any "Mangrove" or significant areas of "ww". Manawatū-Whanganui does not contain "Mangrove", "sd" or "sc". Northland does not contain any significant areas of "ww".

Appendix C: Comparison of the Key Features of Landsat 4-5 (TM) and Landsat 8 (OLI/TIRS) Data

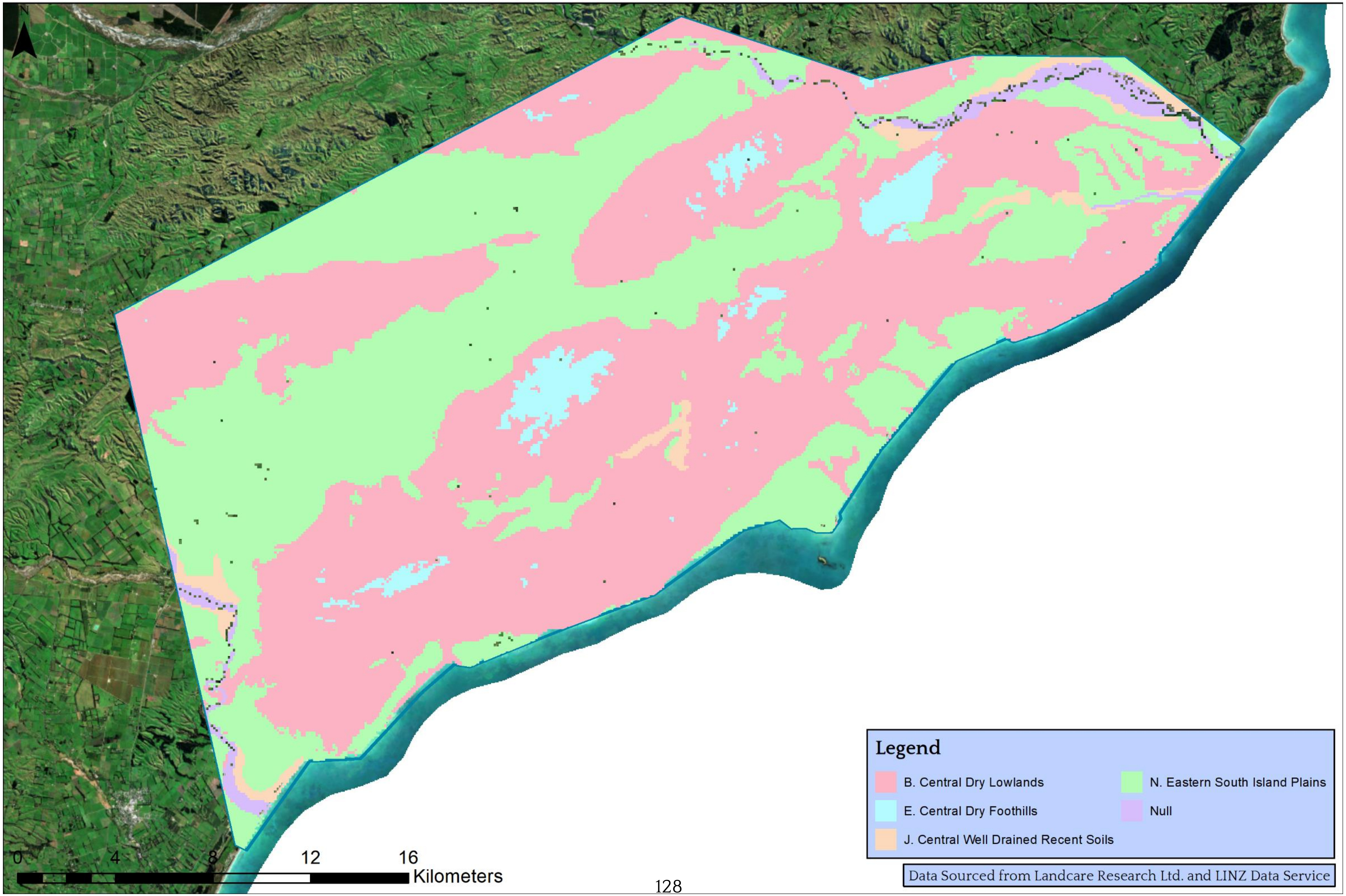
Landsat 4-5 (TM) 1975-2013	Temporal 16 Days	Radiometric Resolution 8 bits	Landsat 8 (OLI/TIRS) 2013-Present	Temporal 16 Days	Radiometric Resolution 12 bits
Band Name	**Spectral Range (µm)**	**Spatial Resolution (m)**	**Band Name**	**Spectral Range (µm)**	**Spatial Resolution (m)**
			Band 1-Ultraviolet	0.43-0.45	30
Band 1-Blue	0.45-0.52	30	Band 2-Blue	0.45-0.51	30
Band 2-Green	0.52-0.60	30	Band 3-Green	0.53-0.59	30
Band 3-Red	0.63-0.69	30	Band 4-Red	0.64-0.67	30
Band 4-NIR	0.76-0.90	30	Band 5-NIR	0.85-0.88	30
Band 5-SWIR1	1.55-1.75	30	Band 6-SWIR1	1.57-1.65	30
Band 7-SWIR2	2.08-2.35	30	Band 7-SWIR2	2.11-2.29	30
-	-	-	Band 8-Panchromatic	0.50-0.68	30
-	-	-	Band 9-Cirrus	1.36-1.38	30
Band 6-TIR	10.40-12.50	120	Band 10-TIR1	10.60-11.19	100
			Band 11-TIR2	11.50-12.51	100

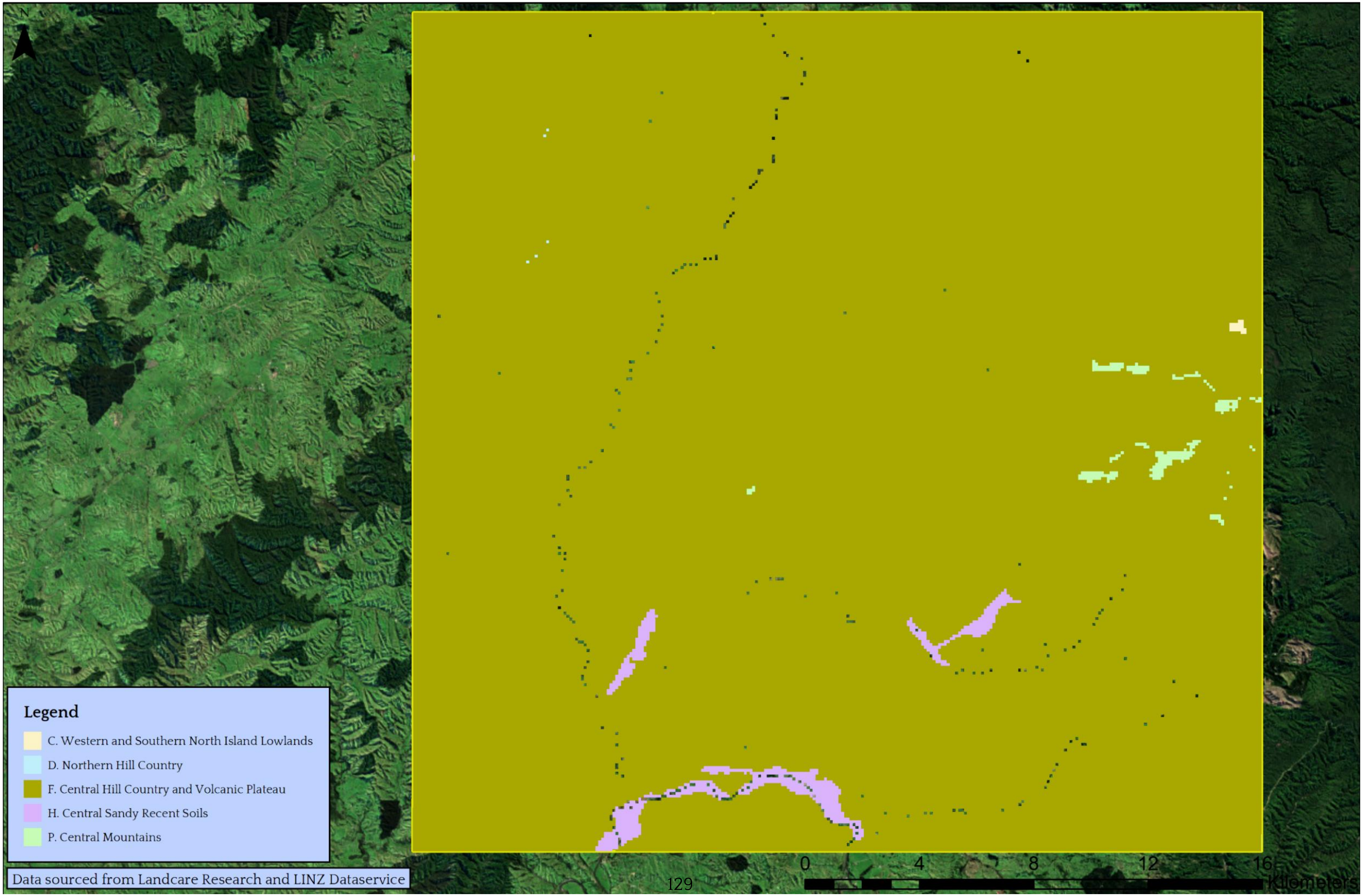
Note. Modified from Phiri and Morgenroth. 2017.

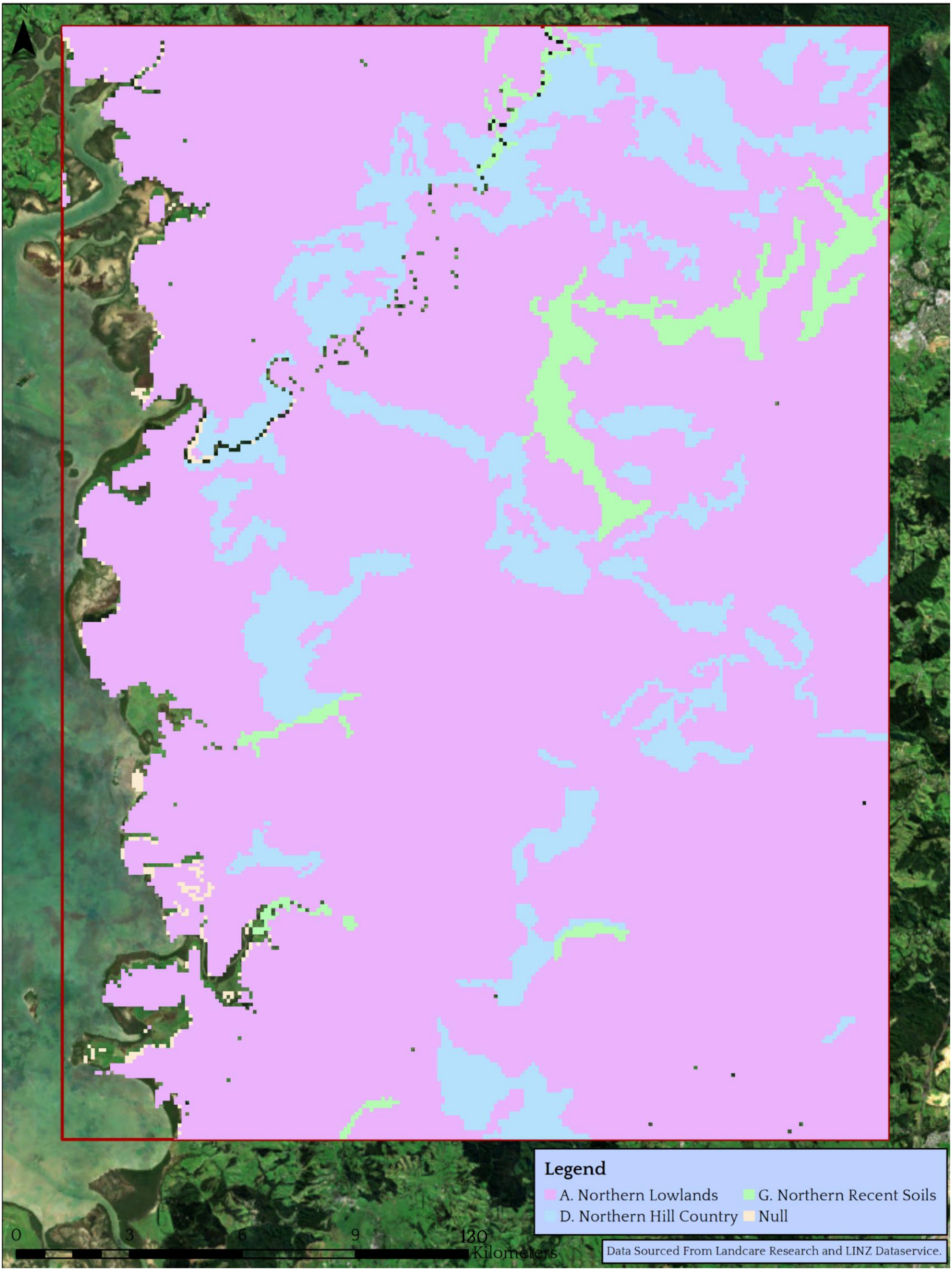
Name	Use	Formula	Source(s)
Wide Dynamic Range Vegetation Index (WDRVI)	Overview of reflectance differences in vegetative land cover. Increased confidence in ground truth boundary drawing.	(0.1*Near Infrared-Red)/(0.1*Near Infrared+Red)	IDB – Index DataBase, (n.d.) Xue & Su, (2017). Gitelson, (2004).
Soil Background Line (SBL)	Discrimination of continuous class pixels from discrete and sparse class pixels in ground truth boundary drawing.	Near Infrared-(2.4*Red)	IDB - Index DataBase. (n.d.). Xue & Su, (2017). Richardson & Weigand, (1977).
Visible Atmospherically Resistant Index (VARI)	Discrimination of edges of different vegetation classes at high contrast.	(Green - Red)/ (Green + Red - Blue)	ESRI Indices Gallery, (2020). Xue & Su, (2017). Gitelson et al. (2002).
Soil Adjusted Vegetation Index (SAVI)	Pre-classification masking of water and bare ground.	((Near Infrared-Red)/(Near Infrared+Red+L))*(1*L)	ESRI Indices Gallery, (2020). Xue & Su, (2017). Huete, (1988)

Note. Spectral range of all band designations given in formulae can be found in Appendix C.

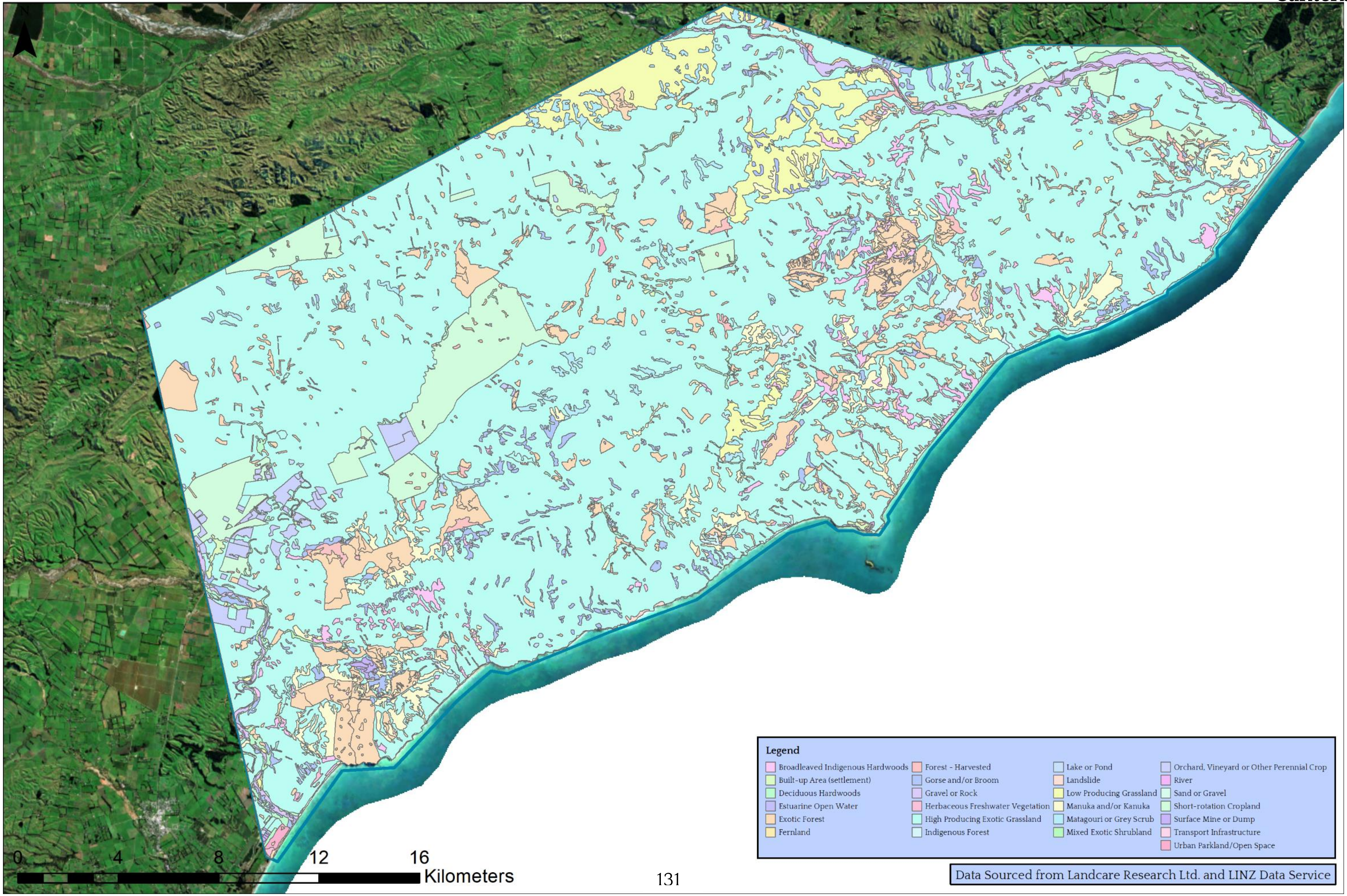
Canterbury

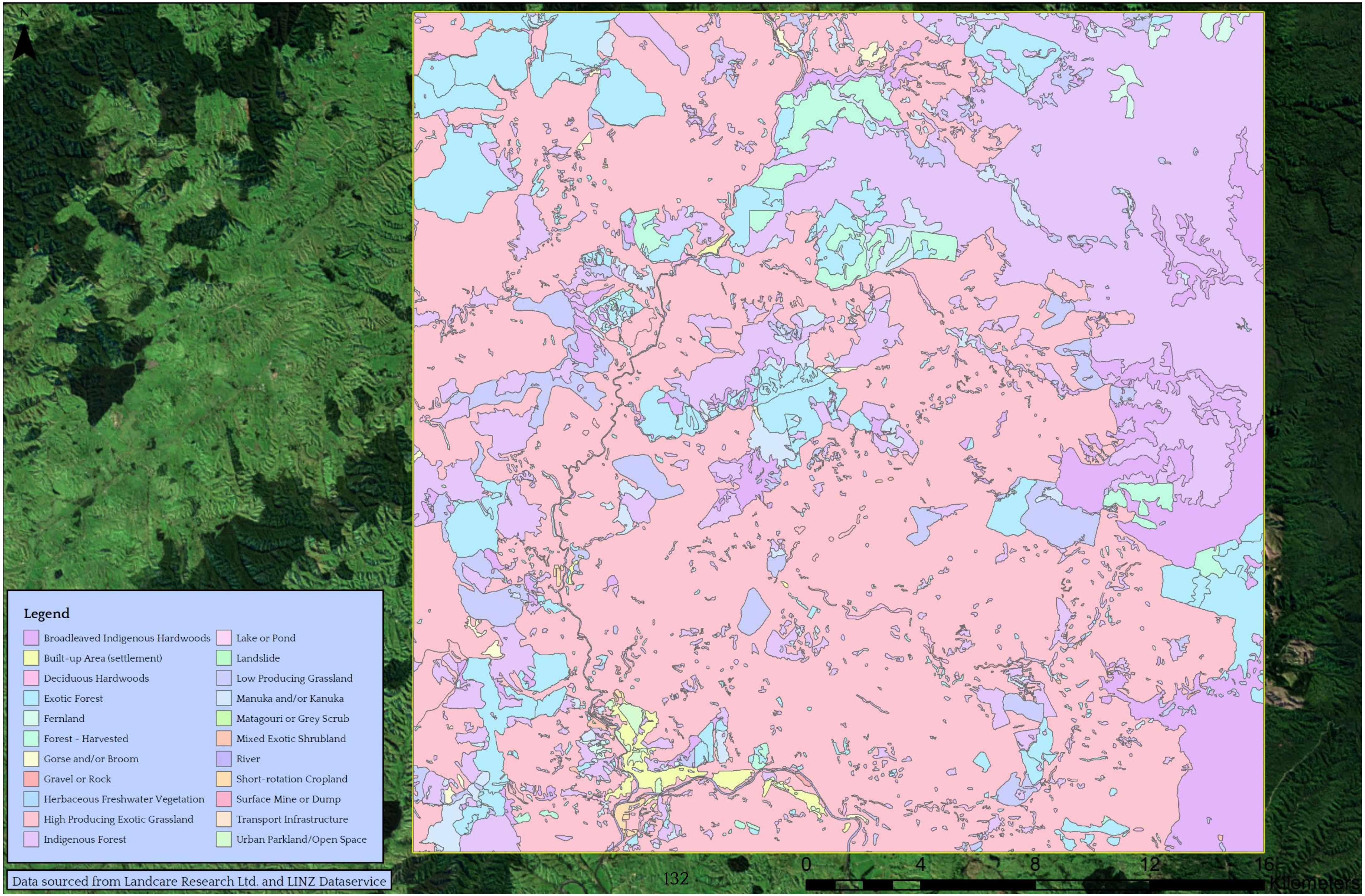

N
0
4
8
12
16
Kilometers
Legend
B. Central Dry Lowlands
E. Central Dry Foothills
J. Central Well Drained Recent Soils
N. Eastern South Island Plains
Null
Data Sourced from Landcare Research Ltd. and LINZ Data Service

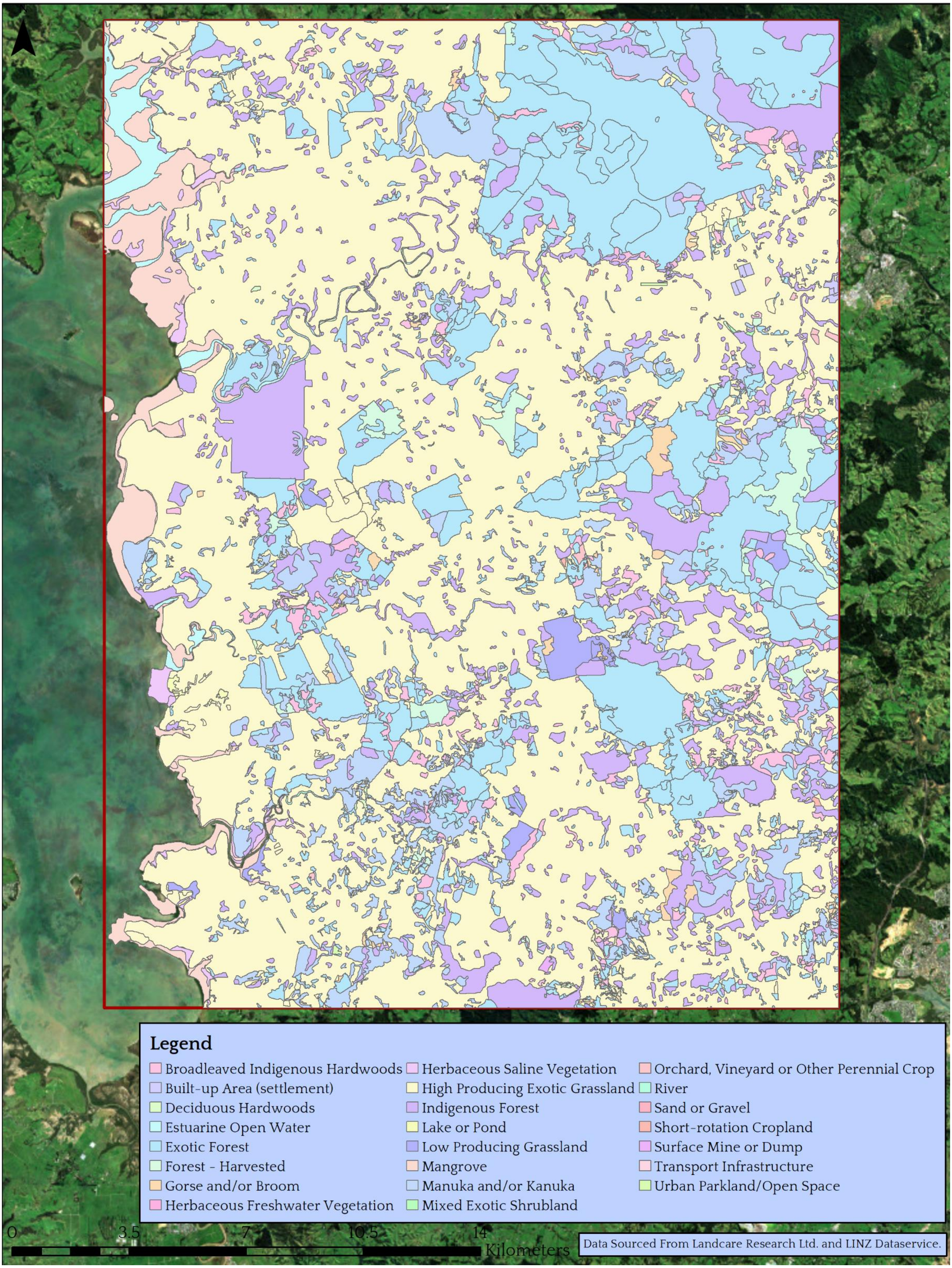

N
Legend
C. Western and Southern North Island Lowlands
D. Northern Hill Country
F. Central Hill Country and Volcanic Plateau
H. Central Sandy Recent Soils
P. Central Mountains
Data sourced from Landcare Research and LINZ Dataservice
0
4
8
12
16
Kilometers


N
Legend
A. Northern Lowlands
G. Northern Recent Soils
D. Northern Hill Country
Null
0
3
6
9
12
Kilometers
Data Sourced From Landcare Research and LINZ Dataservice.

Canterbury


Legend
Broadleaved Indigenous Hardwoods
Built-up Area (settlement)
Deciduous Hardwoods
Estuarine Open Water
Exotic Forest
Fernland
Forest - Harvested
Gorse and/or Broom
Gravel or Rock
Herbaceous Freshwater Vegetation
High Producing Exotic Grassland
Indigenous Forest
Lake or Pond
Landslide
Low Producing Grassland
Manuka and/or Kanuka
Matagouri or Grey Scrub
Mixed Exotic Shrubland
Orchard, Vineyard or Other Perennial Crop
River
Sand or Gravel
Short-rotation Cropland
Surface Mine or Dump
Transport Infrastructure
Urban Parkland/Open Space
0 4 8 12 16
Kilometers
Data Sourced from Landcare Research Ltd. and LINZ Data Service


N
Legend
Broadleaved Indigenous Hardwoods
Built-up Area (settlement)
Deciduous Hardwoods
Exotic Forest
Fernland
Forest - Harvested
Gorse and/or Broom
Gravel or Rock
Herbaceous Freshwater Vegetation
High Producing Exotic Grassland
Indigenous Forest
Lake or Pond
Landslide
Low Producing Grassland
Manuka and/or Kanuka
Matagouri or Grey Scrub
Mixed Exotic Shrubland
River
Short-rotation Cropland
Surface Mine or Dump
Transport Infrastructure
Urban Parkland/Open Space
Data sourced from Landcare Research Ltd. and LINZ Dataservice
0
4
8
12
16
Kilometers


Legend
Broadleaved Indigenous Hardwoods
Built-up Area (settlement)
Deciduous Hardwoods
Estuarine Open Water
Exotic Forest
Forest - Harvested
Gorse and/or Broom
Herbaceous Freshwater Vegetation
Herbaceous Saline Vegetation
High Producing Exotic Grassland
Indigenous Forest
Lake or Pond
Low Producing Grassland
Mangrove
Manuka and/or Kanuka
Mixed Exotic Shrubland
Orchard, Vineyard or Other Perennial Crop
River
Sand or Gravel
Short-rotation Cropland
Surface Mine or Dump
Transport Infrastructure
Urban Parkland/Open Space
0
3.5
7
10.5
14
Kilometers
Data Sourced From Landcare Research Ltd. and LINZ Dataservice.